DOCUMENT RESUME

ED 342 815 TM 017 996

TITLE	Reauthorization of the Office of Educational Research and Improvement Act. Hearing on Examining the Need for a Nationwide Testing System Focusing on Recommendations To Establish National Student Performance Standards before the Subcommittee on Education, Arts, and Humanities of the Committee on Labor and Human Resources. United States Senate, One Hundred Second Congress, First Session (March 7 and June 13, 1991).
INSTITUTION	Congress of the U.S., Washington, D.C. Senate Committee on Labor and Human Resources.
REPORT NO	ISBN-0-16-036815-4; Senate-Hrg-102-256
PUB DATE	91
NOTE	292p.
AVAILABLE FROM	U.S. Government Printing Office, Superintendent of Documents, Congressional Sales Office, Washington, DC 20402.
PUB TYPE	Legal/Legislative/Regulatory Materials (090)
EDRS PRICE	MF01/PC12 Plus Postage.
DESCRIPTORS	Academic Achievement; *Achievement Tests; Comparative Analysis; *Educational Assessment; Educational Legislation; Elementary Secondary Education; Federal Government; *Federal Legislation; Government Role; Hearings; *National Competency Tests; National Programs; School Districts; *Standardized Tests
IDENTIFIERS	Congress 102nd; *Office of Educational Research and Improvement; *Reauthorization Legislation

ABSTRACT

The first of two hearings held in the winter and spring of 1991 to consider the reauthorization of the Office of Educational Research and Improvement focused on the question of national testing. The proposed legislation stipulates that a national test would measure an individual's educational achievement, give the local education agency an indication of how its students compare with others, and highlight strengths and weaknesses in the educational attainment of students on a district-by-district basis. Advocates and opponents of a national test and testing system addressed the hearings. Speeches and prepared statements were presented by: Senator Nancy Kassebaum of Kansas; Governor Roy Romer of Colorado; Jack Anderson, school superintendent; Maureen Daniels, elementary teacher, and the senior officers of a range of education related organizations including: William H. Kolberg, National Alliance of Business; Laura Resnick, Learning Research and Development Center, University of Pittsburgh; Gordon Ambach, Council of Chief State School Officers; Albert Shanker, American Federation of Teachers; and Michael E. Melody, College and Test Publishing, Houghton Mifflin Company. At the second session, during which broader issues relating to educational research and improvement were examined, witnesses included Senator Kassebaum, joined by senators Orrin G. Hatch (Utah), James H. Jeffords (Vermont), and Dan Coats (Indiana). Also testifying were Secretary of Education Lamar Alexander and the following nongovernment witnesses: Jeri Nowakowski, North Carolina Regional Educational Development and Research; Arthur E. Wise, National Council for Accreditation of Teacher Education; Nathaniel M. Semple, Committee for Economic Development; Boyd W. Boehlje, National Assessment Governing Board; and Daniel M. Stewart, The College Board. Prepared statements and other documentation are included. (SLD)

IR

S. HRG. 102-256

REAUTHORIZATION OF THE OFFICE OF EDUCATIONAL RESEARCH AND IMPROVEMENT ACT

ED342815

HEARING

BEFORE THE

SUBCOMMITTEE ON EDUCATION, ARTS AND HUMANITIES

OF THE

COMMITTEE ON LABOR AND HUMAN RESOURCES UNITED STATES SENATE

ONE HUNDRED SECOND CONGRESS

FIRST SESSION

ON

EXAMINING THE NEED FOR A NATIONWIDE TESTING SYSTEM, FOCUSING ON RECOMMENDATIONS TO ESTABLISH NATIONAL STUDENT PERFORMANCE STANDARDS

MARCH 7 AND JUNE 13, 1991

Printed for the use of the Committee on Labor and Human Resources

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
[ ] This document has been reproduced as received from the person or organization originating it
[ ] Minor changes have been made to improve reproduction quality
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

42-921 ▪▪

U.S. GOVERNMENT PRINTING OFFICE
WASHINGTON : 1991

For sale by the U.S. Government Printing Office
Superintendent of Documents, Congressional Sales Office, Washington, DC 20402
ISBN 0-16-036815-4

BEST COPY AVAILABLE

COMMITTEE ON LABOR AND HUMAN RESOURCES

EDWARD M. KENNEDY, Massachusetts, *Chairman*

CLAIBORNE PELL, Rhode Island
HOWARD M. METZENBAUM, Ohio
CHRISTOPHER J. DODD, Connecticut
PAUL SIMON, Illinois
TOM HARKIN, Iowa
BROCK ADAMS, Washington
BARBARA A. MIKULSKI, Maryland
JEFF BINGAMAN, New Mexico
PAUL D. WELLSTONE, Minnesota

ORRIN G. HATCH, Utah
NANCY LANDON KASSEBAUM, Kansas
JAMES M. JEFFORDS, Vermont
DAN COATS, Indiana
STROM THURMOND, South Carolina
DAVE DURENBERGER, Minnesota
THAD COCHRAN, Mississippi

NICK LITTLEFIELD, *Staff Director and Chief Counsel*
KRISTINE A. IVERSON, *Minority Staff Director*

SUBCOMMITTEE ON EDUCATION, ARTS AND HUMANITIES

CLAIBORNE PELL, Rhode Island, *Chairman*

HOWARD M. METZENBAUM, Ohio
CHRISTOPHER J. DODD, Connecticut
PAUL SIMON, Illinois
BARBARA A. MIKULSKI, Maryland
JEFF BINGAMAN, New Mexico
EDWARD M. KENNEDY, Massachusetts
PAUL D. WELLSTONE, Minnesota

NANCY LANDON KASSEBAUM, Kansas
THAD COCHRAN, Mississippi
ORRIN G. HATCH, Utah
JAMES M. JEFFORDS, Vermont
STROM THURMOND, South Carolina
DAN COATS, Indiana

DAVID V. EVANS, *Staff Director*
SUSAN K. HATTAN, *Minority Staff Director*

CONTENTS

STATEMENTS

THURSDAY, MARCH 7, 1991

	Page
Kolberg, William H., president and CEO, National Alliance of Business, Washington,DC; Lauren Resnick, director, Learning Research and Development Center, University of Pittsburgh, Pittsburgh, PA; Gordon Ambach, executive director, Council of Chief State School Officers, Washington, DC; and Jack Anderson, superintendent of schools, East Ramapo Central School District, Spring Valley, NY, on behalf of American Association of School Administrators	4
Prepared Statements of:	
Mr. Kolberg	6
Ms. Resnick (with an attachment)	14
Mr. Ambach	33
Mr. Anderson	37
Kassebaum, Hon. Nancy Landon, a U.S. Senator from the State of Kansas, prepared statement	49
DiPatri, Richard A., vice president, Educate America, Inc., Morristown, NJ; Monty Neill, associate director, FairTest (National Association for Fair and Open Testing), Cambridge, MA; Albert Shanker, president, American Federation of Teachers, Washington, DC, and C.L. Hutchins, executive director, Mid-Continent Regional Educational Laboratory, Aurora, CO	57
Prepared Statements of:	
Mr. DiPatri	59
Mr. Neill	64
Mr. Shanker	73
Mr. Hutchins	77
Tucker, Marc S., president, National Center on Education and the Economy, Rochester, NY; Maureen Daniels, 6th grade teacher, Lynbrook Elementary School, Fairfax, VA, on behalf of National Education Association; and Burton W. Faldet, president, Test Consultants, Ltd., Association of American Publishers, Washington, DC; accompanied by Michael E. Melody, senior vice president, College and Test Publishing, Houghton Mifflin Company	90
Prepared Statements of:	
Mr. Tucker	102
Ms. Daniels	108
Mr. Faldet (with an attachment)	112
Romer, Hon. Roy, Governor of Colorado, prepared statement	134
Melody, Michael E., senior vice president, College and Test Publishing, on behalf of Houghton Mifflin Company, Boston, MA, prepared statement and additional materials	136
Neill, Monty, associate director, National Center for Fair & Open Testing (FairTest), Cambridge, MA, "Open Letter to Congress, Bush Administration, the Governors on NAGB and NAEP Expansion	183

ADDITIONAL MATERIAL

Articles, publications, letters, etc.	
An article entitled "Monitoring and Achieving the National Education Goals"	87
An article entitled "America's Choice: high skills or low wages!"	92
Letters to Senator Pell from:	
Veliotes, Nicholas A., president, Association of American Publishers, Inc., Washington, DC, dated March 22, 1991	179

Page

Letters to Senator Pell from—Continued
Sheehan, Sister Lourdes, Secretary of Education, United States Catholic Conference, Washington, DC, dated March 21, 1991 182

THURSDAY, JUNE 13, 1991

Alexander, Hon. Lamar, Secretary, U.S. Department of Education, Washington, DC, accompanied by Bruno Manno, Acting Assistant Secretary, Office of Educational Research and Improvement 190
Prepared statement of:
Secretary Alexander 196
Kassebaum, Hon. Nancy Landon, a U.S. Senator from the State of Kansas, prepared statement 203
Jeffords, Hon. James M., a U.S. Senator from the State of Vermont, prepared statement 207
Coats, Hon. Dan, a U.S. Senator from the State of Indiana, prepared statement 207
Hatch, Hon. Orrin G., a U.S. Senator from the State of Utah, prepared statement 208
Nowakowski, Jeri, executive director, North Central Regional Educational Development and Research; Arthur E. Wise, president, National Council for Accreditation of Teacher Education, Washington, DC; Nathaniel M. Semple, vice president and secretary, Committee for Economic Development, Washington, DC 212
Prepared statements of:
Ms. Nowakowski 214
Mr. Wise 218
Mr. Semple (with an attachment) 226
Boehlje, Boyd W., board member, National Assessment Governing Board, Washington, DC; and Donald M. Stewart, president, The College Board, Washington, DC 236
Prepared statements of:
Mr. Boehlje (with attachments) 238
Mr. Stewart 253
Research and the Renewal of Education, submitted by Michael A. Resnick, associate executive director, The National Academy of Education, prepared statement 261
National School Boards Association, prepared statement 264
Kolberg, William H., president, National Alliance of Business, prepared statement 277

ADDITIONAL MATERIAL

Articles, publications, letters, etc.
Stewart, Donald M., president, The College Board, Washington, DC, letter dated June 21, 1991 279

REAUTHORIZATION OF THE OFFICE OF EDUCATIONAL RESEARCH AND IMPROVEMENT ACT

THURSDAY, MARCH 7, 1991

U.S. SENATE,
SUBCOMMITTEE ON EDUCATION, ARTS AND HUMANITIES, OF THE
COMMITTEE ON LABOR AND HUMAN RESOURCES,
Washington, DC.

The subcommittee met, pursuant to notice, at 10:04 a.m., in room SD-430, Dirksen Senate Office Building, Senator Claiborne Pell (chairman of the subcommittee) presiding.

Present: Senators Pell, Bingaman, Wellstone, Kassebaum, and Hatch.

OPENING STATEMENT OF SENATOR PELL

Senator PELL. The Subcommittee on Education, Arts and Humanities will come to order. This hearing marks the first hearing on the reauthorization of the Office of Education Research and Improvement, the so-called OERI. There are many issues with which this office is concerned and will be concerned in the future, but none more interesting, more crucial and, perhaps, more controversial than the question of a national test. I know I have been working on this for a good many years.

Because discussion and debate on a national test is so prevalent in education circles today, we thought it would be a good idea to put the issue front and center in our deliberations on OERI.

My own interest in this goes back, as I mentioned earlier, more than 24 years when, in 1967, Senator John Sherman Cooper and I joined forces to introduce the Quality in Education Act. That legislation was to devise a method by which we might be able to compare secondary school education on a district-by-district basis throughout our Nation.

In the aftermath of the proposals set forth in "The Nation's Report Card," in 1987, I resurrected that idea, modified it, and introduced legislation that authorized the Secretary of Education to formulate the Optional Test of Academic Excellence. That legislation became law in 1988. My understanding, though, is that little work, alas, has been done in the department with respect to the Optional Test.

In light of the current national debate and because of the department's inactivity in this area, I will soon introduce legislation that would mandate, i.e., require the Secretary of Education to approve a test or series of tests that would serve as the national test of academic excellence.

Unlike my previous legislation, there is no stipulation that this test would be a voluntary one. We may reach the conclusion that it should be voluntary during the debate on this legislation. But I thought it best from the outset not to set such a limited boundary in legislative language.

My idea is that we should have a national test, not too different from the New York Regents Test. Its purpose would be three-fold: It would measure an individual's educational achievement; it would give the local educational agency an indication of how its students compared with others; and it would point out both strengths and weaknesses in the educational attainment of students on a district-by-district basis.

It would also award a certificate to each student who passed the exam and in that way would help identify talented students would not otherwise be recognized. They could take this certificate to a college or to the workplace as an indication of their educational achievement.

If enacted, the test would inevitably raise questions regarding a national curriculum. Those are questions that must be addressed as part of the debate on a national test. In some ways, we already have the elements of a national curriculum because of the dominance of a few States in selection of textbooks. The requirements set forth by those States tend to establish a floor that all States adhere to because other options are not available.

The question therefore may not be should we have a national curriculum, for indeed the basic elements may already be there. The real question may be what elements do we want to be part of the national curriculum.

I look forward to the testimony we have today, and would suggest the ranking Republican on the full committee may have a statement.

OPENING STATEMENT OF SENATOR HATCH

Senator HATCH. Well, thank you, Mr. Chairman.

The topic we are addressing today is a significant one. Testing is an issue that we must understand clearly if we are to make proper decisions in the future about educational programs and about assessment of our progress. All realize the need to know how effectively our children are taught and how effectively our tax dollars are spent.

I hope we approach this issue with the necessary caution. Since testing is clearly one of the methods being relied upon more frequently as a measure of these outcomes, it behooves each member of this committee to understand just what information testing can and cannot give us. Tests alone cannot adequately assess our educational system. There are other desirable results that need to be taken into consideration, such as the ability of students to accept responsibility, to work well with others regardless of race, religion, physical and mental limitations, and the ability to make good decisions which are morally right.

Many of these outcomes cannot be tested with current testing instruments. We also need to understand the implications of having comparative data that can be used for a variety of purposes both

good and bad. For example, test results can be used to evaluate teachers, principals, and superintendents; or they can affect property values when realtors share data about schools with high test scores and businesses make decisions about where to relocate based on test scores.

It is also imperative that we understand the strengths and weaknesses of the various types of tests and testing programs. In our fervor to measure outcomes and assess progress, we want to make sure that we do not test without first deciding the purpose of testing and then choosing or developing a test that meets that particular purpose.

Utah has embarked on a plan to measure progress by administering the Stanford Achievement Test to every student in the 5th, 8th, and 11th grades. The Utah legislature passed and funded this requirement a year ago and implemented it last fall. At the present time, the test results have been published, but there has not been enough time to see whether or not tests will influence administrative practices, teaching methods, or citizen reactions.

Now, I intend to follow this closely to determine the impact of such mandated State testing programs. Similar testing programs have been implemented in other States, and their experience with such programs can help us determine the efficacy of testing as an assessment tool.

I look forward to reviewing the testimony today and want to thank all of the witnesses being here this morning. I will also particularly welcome the views of Lamar Alexander, who I hope will be our next Secretary of Education without too much further unnecessary delay. And I think any delay beyond a day is unnecessary.

His judgment and expertise will be valuable on the key question of education policy. I am very disappointed that this committee has taken so long to resolve the problems with regard to the confirmation of Lamar Alexander.

I don't know how we could have had a better nominee for education at this time, yet we keep playing around with that nomination as though it is just some inconsequential nomination. We have now had a rudderless Education Department for almost 3 months. Frankly, we can resolve these problems if we will just move as a committee and get them resolved.

I think that you are always going to have the politicians down there in Tennessee resurrecting little problems the rest of our lives if we keep paying attention to them. Frankly, every major question has been answered, answered well, answered with honesty and decorum, and with good logic and analysis. I think it is time to stop the syncopated smear shuffle. Frankly, I am getting kind of sick of it.

I have been promised now by this committee for 3 weeks that they would put this man out. They basically, told us he would be out by Tuesday. And Tuesday has come and gone.

I am not criticizing, this subcommittee, because I know that Senator Pell admires and respects Lamar Alexander and has been very supportive of him. I want to pay tribute to my chairman on this committee and the ranking member. They have worked very hard

to see that education is pushed forward. I just don't know anybody that has given more time to it than Senator Pell.

I am talking about the committee as a whole, and I am using this forum to just vent my spleen because I think it's time for us to get this job done. I hope we can get Lamar Alexander out before the end of business today. If we don't, it's another week that education has gone rudderless at a time when we have the No. 1 person in the governorship of the 50 States who pushed education programs like nobody else did, who could help this education president become the education president that he wants to be. And I think it's time to do it. It's time to put away our differences, time to put away politics and time to put that man out of the committee and onto the floor and into office so that we can move ahead. And I think you will have a person who is innovative, creative, and hard-working with good judgment in that job if we will make that decision.

I would like again to thank Senator Pell for his help thus far in trying to push this nomination forward.

So, thank you, Mr. Chairman. I appreciate it.

Senator PELL. Did you have an opening statement as ranking member?

Senator KASSEBAUM. No. I will make any comments later. I look forward to hearing the panel.

Senator PELL. Mr. Kolberg, welcome.

STATEMENTS OF WILLIAM H. KOLBERG, PRESIDENT AND CEO, NATIONAL ALLIANCE OF BUSINESS, WASHINGTON, DC; LAUREN RESNICK, DIRECTOR, LEARNING RESEARCH AND DEVELOPMENT CENTER, UNIVERSITY OF PITTSBURGH, PITTSBURGH, PA; GORDON AMBACH, EXECUTIVE DIRECTOR, COUNCIL OF CHIEF STATE SCHOOL OFFICERS, WASHINGTON, DC; AND JACK ANDERSON, SUPERINTENDENT OF SCHOOLS, EAST RAMAPO CENTRAL SCHOOL DISTRICT, SPRING VALLEY, NY, ON BEHALF OF AMERICAN ASSOCIATION OF SCHOOL ADMINISTRATORS

Mr. KOLBERG. Mr. Chairman, I am William Kolberg, president of the National Alliance of Business. I appreciate your invitation to appear before this subcommittee today.

The President spoke to all of us last week of the valuable lessons to be learned from the war in the Gulf. That experience shows what this country can do when we prepare to meet a challenge, develop new technology, experiment and refine tactics and strategies, and, in effect, modernize every aspect of knowledge, training, and institutional structures. The results were unparalleled.

We also proved with the moon landing in the 1960's and with this recent Gulf war that when there is consensus in America, we get things done. Currently, we have neither consensus nor standards nor assessment systems on what our students need to know.

We in business recognize that our Nation's failure to educate our young people to world-class standards is a major national economic problem which will be solved only if we have the national will, the new strategies, the superior technology, and the new institutional structures which can lead us to achieving victory in our fight against today's educational mediocrity.

The National Alliance of Business endorses the concept of establishing a framework for a national examination system. I need to emphasize that we support a national system of student assessment, and not a single national test. We see this national system as a varied and cumulative system of assessments over time. It should be a system based on a set of national commonly held standards of performance and competence. Cumulative assessments based on performance will provide multiple opportunities for students to achieve success rather than a single high-stakes moment of possible failure.

U.S. employers today have no reliable gauge of what a new worker knows or doesn't know. The high school diploma is no longer a reliable gauge of competency. Likewise, schools do not know what workforce skill demands are. Fortunately, it seems to me, Mr. Chairman, that we are finally moving and we are seeing some national efforts such as the National Goals Panel and similar local efforts which are beginning to bridge the knowledge gap between curricula learning and the workplace.

The establishment of national goals by the President and the governors was a laudable first step toward defining national expectations for the education enterprise. Now we have to find ways to build the national will to achieve these goals, and we need to be able to measure our progress along the way. Reaching the national goals must be a key national priority. National standards need to be established that set guideposts for what students should know in order to participate as citizens, go to work, receive additional skill training, or attend college.

We in this Nation should have the ability to aggregate the information on individual student performance against these standards. That information should be gathered by school, by school district, by State, and, yes, nationally. Parents should be fully informed about how their children are doing against these national standards. The United States can learn from the experience of other countries. Great Britain, for example, has recently established a national curriculum that all students follow. The teachers in that country have national criteria and assessments to measure student progress.

The U.S. should be willing to challenge a few of the traditions in education, by having a greater national approach to setting standards while still leaving most implementing decisions to State and local levels. It seems to me that we have matured politically enough now to talk about a nationally developed syllabi with criteria and standards against which States and local curricula can be developed.

It is my worry that we are not serious enough yet about making critical changes in education and not committed yet to achieving our national education goals. As an example, the fourth national goal is to be first in the world in math and science education by the year 2000. Today, Mr. Chairman, our students test dead last in the world in math and science.

For us to get from last to first in 9 years will, I think, require a national commitment similar to the one that we made to get a man on the moon in the 1960's. I sense that we are still unwilling as a

Nation to make the changes that would be necessary to achieve these policies.

If we are to succeed, the Federal Government must be much more proactive in its leadership. The Federal Government should be providing moon landing-type leadership in support of developing a system of student assessment. The Federal Government has a traditional role in supporting research upon which new knowledge and innovation in education are based.

This role becomes critical in the 1990's, to explore new methods, develop standards, and devise new assessment mechanisms. The Federal Government must lead in the development of knowledge and basic data from which the entire educational system can draw as it strives to achieve our goals.

Mr. Chairman, business has an important role to play as a partner in this endeavor, and all of the major national business organizations in this city are now involved in very significant education reform activities.

I would be happy to answer any questions later on. Thank you again, Mr. Chairman.

Senator PELL. Thank you very much indeed, Mr. Kolberg.

[The prepared statement of Mr. Kolberg follows:]

PREPARED STATEMENT OF MR. KOLBERG

Mr. Chairman, I appreciate your invitation to appear before the subcommittee today to provide a business perspective on education reform and, in particular, about the need to assess our progress toward education excellence in this country.

I am William H. Kolberg, president, of the National Alliance of Business.

While education remains largely a State and local responsibility in this country, we are now viewing education issues in a national context and as a national problem. This concern has led the National Alliance of Business to join in partnership with other national business organizations in the *Business Coalition for Education Reform* to work over the long haul with education and community leaders to help reverse declines in education quality and economic opportunity. (The Business Coalition for Education Reform includes: American Business Conference, Black Business Council, The Business Roundtable, Business-Higher Education Forum, Chamber of Commerce of the United States, Committee for Economic Development, The Conference Board, National Alliance of Business, National Association of Manufacturers, and the U.S. Hispanic Chamber of Commerce). American education quality is more widely, and more often, talked about among business leaders today than I would have imagined possible just 5 years ago.

The American education enterprise, which worked well for most of this century, has not kept pace with the increasing demands of modern society and our internationally competitive economy.

By way of illustration, the futurist Marvin Cetron has said that when the Class of the Year 2000 graduates from high school, the body of knowledge will have doubled four times since 1988. And these future graduates will have been exposed to more information during the year 2000 than their grandparents consumed in a life time.

We are a society whose future relies heavily on the quality of our workforce. We must dramatically improve both how and what we teach for all our youth and equip them with the knowledge and skills for the world they will enter. It is equally clear, given future projections of fewer labor market entrants, that we can no longer afford to lose 30 percent of our youth who at some point drop out of school before high school completion.

Business has experienced, first hand, the results of lower American educational excellence. U.S. education performance has declined, when compared to other industrialized nations. This shows up in the workplace through lower productivity, lower wage growth, and a diminished competitive posture in the world market. This explains, in part, the compelling interest of business in education, although our interest as citizens in the quality of our society goes beyond the bottom line.

The National Commitment to Education

All of us just witnessed the dramatic events of war in the Persian Gulf. There are some valuable lessons from that experience about what this country can do when we anticipate and prepare for a potential crisis in a fast changing environment. We spent years on the development of new technology, on experimenting and refining new tactics, new strategies, and, in effect, modernizing every aspect of knowledge, training, and institutional structures for our national defense to meet new challenges in the changing world—and the results were demonstrated. It shows that if we really put our minds and resources to something, we can be successful. The same level of urgency, commitment, and "national will" must be applied to education to protect the most fundamental strengths of our democratic society.

We in business now see our failure to educate our young people to world class standards as a major national economic problem which will be solved only if we apply the national will, new strategies, superior technology, and new institutional structures which can lead us to a comparable victory by the year 2000.

The industrialized nations who have become our key competitors around the world are those who long ago recognized the importance of education for their economic wellbeing. They each have adopted a national policy and practice for a systematic transition from school to work for all youth, and, with a high level of education skills in their workers, each has been able to organize work more efficiently with greater productivity by cultivating higher skills in front line workers. Their approaches to developing work opportunities through high skills and high expectations draws a sharp contrast to America's current approach.

National Standards and the Assessment of Process

The National Alliance of Business and other organizations endorse the notion of establishing a framework for national student assessments. I need to emphasize the plural *assessments*—not one test. We see this as a varied and cumulative system of assessments over time. And, it is a system based on a set of national standards. Our objective is for every student to leave school with a demonstrated ability to read, write, compute and perform at world class levels in general school subjects, and also to be able to learn, think, and work effectively both in groups and alone.

The concept of assessment which the Alliance supports would facilitate development of assessment systems that provide a variety of ways for students to demonstrate knowledge and skills. *We do not want a single national test, but, rather, commonly held standards of performance and competence, demonstrated through a broadly understood system of diverse assessment techniques.* Students would be assessed over a period of years. Cumulative assessment provides multiple opportunities for success rather than a single high-stakes moment of possible failure. It will enhance opportunities for the undereducated and under motivated to achieve high educational standards at their own pace, because the criteria for the assessments would not vary, regardless of the student's age. Although we advocate national standards of performance and a common assessment system, we also believe that the system should be administered locally, not nationally.

U.S. employers have no reliable gauge of what a new worker knows or doesn't know. Likewise, schools do not know what workforce skill demands are. Fortunately, we are finally seeing national and local efforts to bridge the knowledge gap between curricula learning and the workplace.

The National Education Goals Panel is developing recommendations on standards and assessment systems. The work of the Secretary (of Labor's) Commission on Achieving Necessary Skills (SCANS) will provide an important starting point by identifying the types and levels of skills needed for success in the competitive workforce. The National Center on Education and the Economy is developing a national assessment system which we think has real potential.

Once new assessment instruments are developed, based upon new goals and standards, they will take time to be widely disseminated, used effectively, and accepted. Even when implemented, substantial changes in student performance will not be quickly forthcoming. In fact, some of the new efforts may fail. Restructuring, by its very nature, rests on a willingness to experiment. Like businesses that restructure, efforts will not always succeed, but can point to more effective future practices.

Business leaders and educators are already combining efforts to define the standards of competence for entry into the workforce. Connecticut, New York, California, and Vermont, are developing new assessment tools and systems, ones that are so comprehensive in the information they capture that they can enhance student learning and can even form the basis for teacher and administrator assessment sys-

tems. They clarify learning goals, and more thoroughly measure the higher level skills that business and society require.

An assessment system should be viewed as something developmental that you learn from and refine as you do it. in fact, such systems should continually evolve, because the knowledge needed to succeed in life and work will change.

Changing the accountability system in education through assessment will also affect the management and administration of education, curriculum and instruction, and educational professionalism. The development of assessments cannot be separate from the professional development of educators, because it will have an impact on the nature of instruction and curricula content.

Related to the professional development of teachers to manage and implement an assessment system against standards, I know Mr. Chairman, that you have introduced legislation for the professional development and training of teachers—the "National Teachers Act"—and we would welcome the opportunity to testify when you schedule hearings on that subject.

National Leadership and the Federal Role in Education Reform

We still lack many of the tools necessary to achieve broad-based and lasting success in our efforts to reform education. The establishment of national goals by the President and governors was a laudable first step toward defining national expectations for the education enterprise. These help to articulate where we want to go. Now we have to find the ways to build the national will to achieve these goals, and we need to be able to measure our progress along the way. Reaching the national goals must be a key national priority. I recognize the complexities involved, but we must articulate and apply a clear national strategy, have more cohesive leadership, and provide the necessary resources to succeed.

Currently, we have neither a national consensus nor standards, on what students need to know. We are told by teachers that they do not have a clear idea of what they should be teaching differently. We are told by parents that they do not know how their children are doing compared to other students, let alone among other industrialized societies. (In fact, parents are so confused that 75 percent of them believe there is a major education crisis, but 75 percent of them also think that the school their child attends is very good.) And students understand very little about what they will need to know to succeed in contemporary society.

First, national standards need to be established that set guideposts, a framework, and the expectations for what students should know. Standards should define what they need to know—by age—for being a citizen, for going to work, receiving additional skill training, or attending college.

We should have the ability to accumulate information about where individual students are in meeting these broad standards. That information should be gathered by school, by school district, by State, and nationally. Parents should be fully informed about how their children are doing.

The education standards we adopt should, of course, be competitive with other nations, but be designed to meet U.S. needs first. Our workers must be among the finest, if we expect our businesses to be as efficient as our international competitors, many of whom have national educational assessments.

The United States, like Great Britain and other nations, need to address this problem directly. Britain has established a national curriculum that all students follow. It is standardized. Teachers have national criteria and assessments to measure student progress. Britain went at the problem directly by specifying curricula at the beginning, rather than having educational standards and assessments define the changes to curricula at a State and local level. What the British have done would be a revolutionary concept in this country. I use Great Britain to illustrate how strongly other nations are committed to the solution of their education problems.

We should be willing to challenge a few of the traditions in education by having a greater national approach, while still leaving most policy decisions to State and local levels. We've matured politically. We can talk now about national curriculum in this country, based on standards we establish. The important distinction is that this would not be a Federal curriculum. I believe there is a role for federally developed criteria, standards, or syllabi against which State and local curricula can be developed.

It is my worry that we are not serious enough yet about making critical changes in education, and not committed yet to achieving our national education goals. For example, the fourth goal is to be first in the world in math and science education by the year 2000. For us to get there, it will require a national commitment comparable to the challenge of putting a man on the moon in the 1960's. I sense that we are

still unwilling, as a Nation, to think about the kinds of changes that would be necessary in policy, in structure, and in funding to achieve these goals. Without clear national and State leadership, educators are not likely to get behind the efforts with any sense of national mission.

The Federal Government, for its part, must be much more pro-active in its leadership to set the tone, set the agenda, and motivate action. There are several Federal roles which can be played by the administration and the Congress. The Federal Government should be undertaking a "moon landing" leadership role in developing standards for meeting our educational goals, so that there can be no doubt about what States need to do and where children rank in terms of what we have to achieve by the end of this century.

I firmly believe that there is a Federal role in this effort as the lead partner in developing information, establishing standards, and devising assessments.

We are now at a juncture in our efforts to reform education where we recognize that we must act as a Nation on a national challenge, and yet the needed actions must primarily be taken by 50 States and 15,000 local school districts, and, ultimately, in 83,000 school buildings. The Federal Government has historically been a relatively passive force in education and provides less than 10 percent of the funds for education. The challenge is to define a pro-active national leadership role for the Federal Government without displacing or supplanting the traditional State and local responsibilities for education.

In addition to our strongly held belief that the Federal Government must play a critical role in the development and establishment of educational standards and help to design a system of assessments, there are other specific activities appropriate to Federal responsibilities in meeting our national education goals.

National Leadership through Research and Information Development. The Federal Government has a traditional, and accepted, role in supporting research upon which new knowledge and innovation in education are based. This role needs to be expanded in the 90's, as we search for new methods, develop standards and new assessment mechanisms, and to otherwise work toward meeting our ambitious goals. What is needed is a strategic approach to information development and the use of data in relation to national goals. We should carefully, but pro-actively, develop the national capacity to provide the knowledge and basic data from which the entire educational system can draw.

The single largest pot of research money that could be directed to these efforts is under the Education Department's Office of Educational Research and Improvement (OERI). What we should buy is a new strategy for the future education of our youth. The department should be an important catalyst for change, and should underwrite much of the developmental work that is necessary.

My sense is that the research agenda of this office has been to fund a broad range of small, discrete research projects which do not add up to a strategic plan or vision, does not provide information which the States and local schools need for their efforts to restructure education, and is not widely or systematically disseminated when completed. We are missing an opportunity for the Federal Government to provide the leadership in research on education restructuring and national goals.

Early Childhood Development. This is a programmatic area in which Federal leadership has had tremendous impact. It is the best example of filling a gap in the traditional systems of mandatory education. Early childhood development is an important new concept in education, as a strategy of prevention, of which the Head Start program is a part. The Committee for Economic Development (CED) has provided leadership on this issue, and has not only educated the business community about the importance of preschool education and health care, but also has argued convincingly for pursuing a strategy of prevention in public policy. Just yesterday, five Chief Executive Officers of major U.S. corporations testified before the House Budget Committee on full-funding of the Women, Infants, and Children (WIC) program.

We recognize how critical early childhood education can be. It has a direct impact on social skills, educational achievement, and self esteem. We at the Alliance see investments in early childhood programs, like Head 'Start, as an important weapon in the fight against the problems of school dropouts, drug abuse, crime, and teenage pregnancy and for that reason we are on record in support of full funding for Head Start. It has the potential, over the long term, of allowing us to redirect limited Federal dollars that otherwise might have to be spent on "second chance" systems to repair the damage that could have been prevented. I would also argue that we are at a point where the costs could be shared with the States. We know that about 30 States have enacted various types of early childhood programs, 9 of which are designed to supplement Head Start.

Now, with the governors recognizing a national goal related to early childhood education, and with the States having primary responsibility for public education, perhaps we could move to a greater level of shared responsibility in early childhood development. There is precedent in virtually every other program of Federal assistance to education. The closest comparison is the Federal Chapter 1 program, which predominately covers poor children in the early years of elementary school, in which costs are shared with the States.

Build Linkages in the Broad Range of Federal Education Programs. We must take much more care to rationalize how individual programs are linked in a cohesive continuum of education development. Individual Federal education programs must be thought of in relation to each other. For example, Head Start cannot be separated from services under Chapter 1 of the Elementary and Secondary Education Act or we risk losing the gains of one program during the application of another. Each program should build logically on the progress of the others. We still need to emphasize educational services to disadvantaged groups, but the policy must be to build on linking education programs. This also can have an effect in reducing the bureaucracies which have been established over time at the Federal, State, and local levels.

Integrate Health and Human Service Programs with Education. Educators are often the first to identify health or other problems that are preventing children from learning. But, they are often helpless in addressing these problems. Federal and State legislation must put a premium on assuring that health and human service programs are readily available to children. This can be accomplished by insisting that such programs demonstrate how well they are tied into the schools and responding to the problems identified by school teachers, counselors, and administrators.

Establish a Better School to Work Transition. For several years now, various policy studies have focused on the failure of our society to provide school to work assistance to the majority of students who do not go on to college. We are not naive about how complicated this issue is, but it deserves urgent attention by this subcommittee. I understand that several members of the full committee have already expressed interest in working on such a proposal. In my view, it involves not only integrating opportunities for work experience with school to give relevance to classroom learning and to motivate students (as the Europeans do), but also to assist students with the skills for finding meaningful employment. America prepares only a tiny fraction of its non-college bound students for work. Other industrial nations have multi-year career educational programs that prepare students to operate at a professional level in the workplace. The enactment of the "Tech-Prep" or "Two-Plus-Two" program in the vocational education reauthorization last year is an important step in this direction and may serve as a model for a more extensive system of occupational certification. The Alliance intends to develop more detailed ideas on the Federal role in a school to work transition, and we will work closely with the committee during the this session as our work progresses.

Provide a Safety Net for Those Who Would Otherwise Fail. This is an important and traditional role of the Federal Government in education to assure equal opportunity and equity of services in education. We would like to see this role broadened in the way help is provided to school dropouts. There are a variety of existing programs in this area that need to be linked more carefully into a cohesive strategy. All students should be guaranteed the educational attention necessary to gain mastery of a standard set of educational skills by age 16, or as soon as possible thereafter.

Insist on Accountability. One critical lever that the Federal Government has over its investment in education is to carefully structure and insist on accountability. This means not only fiscal accountability, but also accountability for solid results. The Chapter 1 accountability standards are an example of what is needed. We are not prepared today to recommend specific methods to achieve accountability, but we do believe that rewards and consequences should be a part of education program legislation.

The Business Role

As for the business role, the business community must be included as a partner in this endeavor. Business can provide information on competency levels for jobs, so that learning experiences and curricula meet the needs of a dynamic society. Business can provide technical assistance in developing standards, assessment tools, and strategies. Business can work closely to help in training of teachers to apply these

new assessment techniques. Business must also be an advocate for a comprehensive system of assessments and continue to press for systemic change.

Business leaders can be instrumental in keeping education high on the public agenda in their States and communities. They can be strong advocates for the transformation of the schools. They can help raise the sights of educators who, feeling powerless and frustrated, often lose any incentive to press forward. Joint efforts are necessary to address the spectrum of education issues in a coordinated and focused approach. Business leaders must work collaboratively and over the long term with educators as well as community leaders toward common goals.

IN CONCLUSION

Mr. Chairman, this is a complex agenda for change. Despite what seem like insurmountable obstacles, a growing current of public opinion demands change in education. All Americans must play a part. We in business are preparing ourselves to play an important role in achieving significant change and improvement.

Our long term agenda, through the Alliance's Center for Excellence in Education and with our partners in the Business Coalition for Education Reform, is to find and implement more effective ways for business involvement.

This education reform effort requires strong national leadership from the Federal Government in setting the vision and the goals, fostering change, and in ensuring that all the stakeholders carry out their appropriate roles.

As America struggles to improve its education system, so that our society can be a strong and informed democracy and can compete in a global market place, it is essential that we are able to assess, in a meaningful way, the educational progress of our students. We are united in our determination to prepare our children to meet the challenges of the future. An informed, accurate, and fair assessment system is an important tool toward that goal.

I will be happy to answer any questions you may have.

Senator PELL. Dr. Resnick, we would be glad to hear from you.

Ms. RESNICK. Thank you. I am Lauren Resnick. I am director of the Learning Research and Development Center at the University of Pittsburgh, and I come to you today as both a long-term scholar of examinations and testing and recently an advocate. I want to explain how I have come to be an advocate. It comes from studies of testing and examining practice here and abroad, not so much as a technical problem of measuring but as a study of how tests function socially to set standards, to encourage people to work or not to work, and to really move schools along in an achievement agenda.

The conclusion we reached is that the kind of testing that America has works against what we want. The current tests are aimed at routine skills, not at thinking, problem-solving, decision-making, those kinds of abilities that have come to be called the new basics that are needed for functioning in today's restructured workplaces and in civic life.

The current tests discourage work in school instead of encouraging and shaping it. They are deliberately decoupled from the curriculum. They are not designed to be studied for. When teachers teach to the test, they are considered vaguely unethical, and indeed we know from research that we have done that when you teach to the test you raise test scores—the tests we have now—you raise test scores and you actually lower real achievement.

That is partly because these tests don't set clear standards. The grade-level scores don't say anything about what is in the tests. And as soon as too many children start to meet whatever the grade-level standard is, you have to raise the standards so that you can avoid the "Lake Woebegone" effect.

So there is no way for the kids in the bottom half, so to speak, to ever succeed. And those tests that we call criterion reference

barely help because of the secrecy surrounding them. The standards they set remain fuzzy and hidden. People can't work toward them. And this is a system that is both unfair and unproductive.

What we think is needed are not more tests, not even a different test, but examinations. What we mean by the term examinations is something rather simple that every country in the world but us uses, something you study for, something you prepare for, something that your teachers can help you do better on, preferably, with external grading—that is, by some other teacher, not your own—so that the teachers and students can be allies on this same side of the team instead of adversaries.

Here are some criteria for the kind of examining system we think is needed for this country:

A system of high standards, really high ones. We know that you can use standard-setting to raise achievement. We know that from the minimum competency movement that started in the mid-1970's. A number of States put tests into place that were required for graduation. Lots of kids didn't pass them at first; after a while they began to pass them. So we know that standard-setting can work, but now we have to use the strategy for maximum competency for the new basics that we need now.

We are going to need multiple forms of examination for multiple goals: Complex performances that can be rated by judges; portfolios of meaningful work that students collect over time and that juries can evaluate; projects that allow us to watch what students can do with others in workplaces and civic projects, and use those, as well, as part of the assessment system.

We need an examining system that is designed unlike any European system for all and not just for an elite. We can have both equity and excellence if we allow students to cumulate exam credits over time, if we allow for elasticity in how long it takes to meet a standard but not for what the standard is, and if we clearly communicate what the standard is so that everybody, not just the students who come from privileged families, can know now what they are trying to do.

We need shared high standards nationally set but multiple examinations administered under State and local control. This can be done under a system of anchor exams and calibration procedures that we are now studying and working on with the best technical help in the country. This is both politically necessarily and educationally right, because the energy lies in our communities and in our schools.

We need widespread involvement in developing standards and exams, a consultative process that truly reaches down into every community and school house, and we need exams embedded in a systemic school reform program so that they are not just a measure of continuing failure but, instead, a lever for new success.

The kind of action that could be taken right away is to begin to work on a national educational standards framework. This needs to be consultative right down to the communities, as I have said, but it also needs to be informed by the best national and even international work. It doesn't have to wait for formation of a permanent body or standards board. A commission or a national study group, perhaps with a mixture of public and private funding could start

the process. At the same time, I think we need to institute a crash R&D program to develop and validate new examining forms suitable for being studied for and the methods of calibration that we will need for this system. And we need to make sure that schools and teachers are intimately involved as well as the top technical people in the country.

I think we need to treat the problem as a national mobilization. The ideas and talent exist to bring it off, and I think this is the moment to give it a real try.

Senator PELL. Thank you very much indeed.

[The prepared statement of Ms. Resnick (with an attachment) follows:]

Prepared Remarks of

Lauren B. Resnick, Director
Learning Research and Development Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260
412-624-7485

before the

Senate Committee on Labor and Human Resources
Subcomittee on Education, Arts and Humanities
Senator Claiborne Pell, Chairman

March 7, 1991

Thank you for inviting me to speak with you, Mr. Chairman and members of the Committee. I am Director of the Learning Research and Development Center at the University of Pittsburgh, a facility which serves as the U.S. Department of Education's center for the study of learning. And I am a past president of the American Educational Research Association.

Today I come to you as both a scholar and an advocate. For the past 12 years, my colleagues and I have been studying how tests and examinations work, both here and abroad. Our studies have focused on how testing practices affect the entire educational process, and society in general. We have investigated questions such as who uses tests, and for what purposes; how tests and exams can promote learning; how they can interfere; and the roles that tests and exams play in public awareness of, and support for, education.

Last summer, prompted by my work with the Commission on the Skills of the American Workforce, I switched from studying examination systems around the world to actively advocating one for this country. With Marc Tucker, President of the National Center on Education and the Economy, who also is testifying today, I launched the New Standards Project, an effort fundedat startup by private foundations.

The New Standards Project has begun work on an examination system that is based on many of the findings and ideas I will share with you here. These findings and ideas suggest a need for dramatic changes in the nation's testing practices. Let me start by explaining why.

The 'New Basics': Moving From Routine Skills to the Capacity for Thinking

A new vision of education has emerged in the last several years, fueled partly by the needs of a changing economy and partly by recent research on learning. In this view, education must now focus on teaching higher-order abilities such as complex reasoning, decision-making, and the ability to go beyond the routine. These skills are demanded by new technologies that require the user to understand them, and by dispersed management systems that call for more judgment on the part of each employee. Also, conditions of work are now likely to change several times during a person's work life, requiring a strong capacity for adaptive learning.

Thus when employers say that students leaving high school are not prepared for work, they often mean more than that these people lack the so-called old basics of simple reading, writing and arithmetic. Like colleges, today's employers are asking for graduates who have the *new basics* of critical thinking, problem-solving, and learning how to learn.

The education system we have inherited was not designed, by and large, to prepare people for such adaptive functioning. Its goals for most students did not include higher-order abilities. These goals were reserved for the elite — at first in separate schools, and more recently in certain tracks within comprehensive schools. The schooling of most Americans still focuses mainly on routinized learning. Students learn — or sometimes do not learn — the old basics, but have little chance to learn the new basics.

While it is not new to include thinking, problem-solving, and reasoning in some students' schooling, it is a significant new challenge to include it in *everyone's* schooling. It is new to make it a regular aspect of the school program for the entire population — even minorities, even non-English-speakers, even children of the poor. Meeting this challenge will require a reorientation in which an emphasis on *thinking* pervades the whole education environment, from the earliest grades.

Current Testing Practice: Reinforcement for Routine Skills

But when it comes to testing, we have only the tools of the past. Our testing theory and practice are products of a Tayloristic view of learning in which knowledge and skills can be broken into many little bits, each to be acquired and practiced separately, without reference to when or how it might be used. Current standardized tests do not test students' abilities to solve complex problems, or to figure out something on their own. Reading and math tests favor superficial answers not based on real understanding. For example:

- An analysis of the reading sections of the tests most often used in state assessment programs shows that they ask for many bits of information rather than coherent, organized responses. On the tests for 8th to 11th

grade, the average length of the passage to be read is only about 300 words, and after reading it, students have to answer five to eight questions in about six minutes.

- In math, too, there are strings of unrelated questions that students have to answer at a rate of about one per minute.

- Multiple choice items predominate, so that students almost never have to figure things out for themselves, only choose from among precoded answers.

The Old Assumption That 'Native Ability' Counts Most

In another way, too, today's standardized tests are ill-suited to play their part in reforming American schools. They are based on the old assumption that native ability, rather than organized effort, determines the outcomes of schooling.

This assumption shows up in the fact that most standardized tests are intentionally de-coupled from the schooling process. They are not designed to be studied for, or practiced for. And they are meant to be as "curriculum-neutral"as possible — neither strongly depending on, nor influencing, exactly what is taught in the classroom and how it is taught.

The historical roots of this testing philosophy are worth looking at. In the 1920s, American educators become excited about the ways in which a new technology of standardized achievement tests — based on the aptitude tests given to more than a million and a half men for the first time in World War I — might help the schools. The tests that were introduced then did not connect directly to the curriculum of the schools. But that was seen as a major asset, not a problem, for the tests were designed to recognize abilities established at birth. At the time, psychologists believed that nature rather than nurture was the chief determinant of success in life — and that schools would become more efficient if they recognized and accepted native differences.

Education, in that view, could only ratify the advantages conferred at birth. And standardized tests (people thought) could help schools to classify and track students, thus leading to more adaptive programs and lower failure rates. But critics such as Walter Lippmann predicted that the vogue of native-ability testing would weaken the work ethic in schools, and would actually reduce the power of schools to improve students' mobility and help them advance in life.

Today various kinds of standardized tests are now labeled as "school ability" or "achievement" tests, but on the whole they are still very weakly coupled to what is done in the classroom. These tests are given to school children for district, state, or federal purposes one or more times a year, and their results are considered quite important. Meanwhile, even though psychologists no longer give the same the same weight to the effects of "native ability" on later success, there is no countervailing system of testing to reward and encourage real achievement based on hard work within the school program.

How Testing Influences Teaching: The Thermometer Effect

From NAEP to the SAT to the commercial tests that are the backbone of American testing practice, tests are still designed on the theory that they will neither depend on, *nor directly influence*, curriculum and teaching.

But it doesn't work that way in practice. There is plenty of evidence that when high attention and high stakes are placed on test results, educational effort is directed accordingly, and the test scores in fact will rise. Unfortunately, there is just as much evidence that real educational achievement is likely to *decline* as a result. This happens because teaching effort is usually focused narrowly on the kinds of items (and even the forms of response) that are used in the tests. In turn, this drives out attention to the new basics of critical thinking, problem-solving, and real-world competencies.

Researchers have studied this process in Texas, for instance, where teachers were given suggested instructional strategies for each of the objectives on the state assessment. The strategies were specific to test item forms, and promised teachers high scores for their students. Commercial programs to help students learn test-taking skills also were found to be item-form-specific, and made similar promises.

Investigators in Texas and elsewhere find that under these conditions, a narrower range of skills is taught, and ways of exercising skills that might be equally important are ignored in favor of drilling students on the forms of response they'll encounter on the tests.

This is unfortunate, because today's tests are designed as indirect statistical *indicators* of competence — not *direct observations of competence in action*. Testing philosophy is similar, in this way, to the philosophy of using simple indicators rather than more complex direct measures in the physical sciences.

For example, a thermometer on a wall is an indirect "indicator" of the amount of ambient heat in a room. We look at the height of mercury in a confined column, and read a simple score from it: a temperature reading, in degrees. This indirect measurement works well, largely because it has no effect on the *producers* of heat: the molecules of air in motion. The molecules of air are not "motivated" to produce a certain temperature.

But teachers and school principals *are* motivated to produce good test scores. When the indicator in their school system falls below an acceptable level, they work to raise it. They work, moreover, on the indicator itself. It is as if, when a room had become too cold, the molecules of air converged on the thermometer and danced around it ... heating up the air at just that point, but letting the rest of the room grow even colder.

Some Conclusions for Examination Design

Whether we like it or not, then, what is tested and what is taught are intimately related — at least when high stakes are attached to the test. There is no serious possibility of creating public accountability systems that will not influence what is taught and how it is taught in the schools.

This means we need to start thinking of *every* test or assessment as an instrument that will profoundly affect education. From findings like those mentioned above, my colleagues and I have developed the following principles for examination design:

- *You get what you assess.* Teachers will "teach to the test," if the test matters in their own or their students' lives.
- *You don't get what you don't assess.* What does not appear on tests tends to disappear from the classroom.
- *It makes sense to build assessments toward which you want teachers to teach.* They must be designed as such; it doesn't work to use the old tests for this new purpose.
- *Assessments should be as complex as the skills they hope to measure and reward.* "Indicators" will no longer do — because what is left out of the tests will be left out of the education process.

Examination Systems: The Key to High-Performance Education

What all of this means is that America needs an *examination system*, not more testing. By "examinations" I mean something simple, something the world knows about and uses, but only America resists: direct observations of performances on a set of knowledge and skills that has been publicly agreed to. They must be performances that one can prepare for, in organized courses of personal study. Exams are something you study for, where your effort pays off in recognized achievement.

The United States is the most overtested and underexamined nation in the world. In the course of their schooling, American students take many more standardized tests than students in any other country. But very few Americans ever prepare for a true external examination — one that is graded by someone other than his or her own teacher; one that has personal consequences.

Teachers generally are not supposed to directly prepare students for tests such as standardized achievement and state competency tests, the National Assessment of Education Progress, or college entrance tests. When they do, the public is skeptical that any real learning has taken place, and questions of unethical practice are sometimes raised. Students likewise are not supposed to study for the tests, except in cram courses that no one believes have lasting value.

As a result of this, the American education system does not clearly reward academic effort on the part of either teachers or students. Unlike athletic coaches, who know exactly what kind of contest faces their teams, teachers are not really in a position to train their students for high performance. Indeed, teachers and students often become adversaries instead of allies in a joint effort to meet a known standard.

Furthermore, in the current system of testing, students are not held to a clear standard of achievement that matters to them. At least for the non-college- bound, the tests that they take so frequently do not make much difference in their personal lives. Even grades in coursework do not clearly matter — because employers, doubting that school grades reflect a real standard of performance, rarely check an applicant's school records. At most, they ask whether the applicant has a diploma. And even the diploma is rarely tied to any real standard of achievement or mastery; in many cases, it's mainly a mark of the student's willingness to keep coming to school.

Thus it is not at all clear to many in the education system that organized effort can contribute to their lives. Small wonder, then, that this effort is so often absent.

An examination system can fundamentally change the situation. At the heart of such a system must be a series of examinations that students can explicitly prepare for, which promote and reward higher-order abilities and hard work, and for which teachers can serve as students' coaches, mentors and allies. The system of examinations that we propose would have several features, as described in the following sections.

Six Criteria for a New Examination System

A new examination system begins with the setting of standards, and proceeds to considerations of actual exam design and methods of implementing the system. I believe that we need six elements in the system:

1) Examinations to high standards

We know that setting clear standards and expecting students to meet them can produce results. Some states tried this strategy in the mid-1970s, when they established minimum competency tests. Passing the tests was required for high school graduation. Students were allowed to retake them until they passed. Legal challenges — based on equity concerns — delayed the implementation of the tests, but once they were in place, the rates of test-passing rose as students and teachers came to understand their content. Setting a standard worked!

Now we have to use this strategy to set standards for *maximum competency*. We want to be sure students are learning what it takes to succeed in the real world — not just what it takes to get high scores on tests. This includes:

- Being able to think critically and solve complex problems.
- Being able to *define* problems out of the "mess" of everyday life in the workplace or home.
- Planning ahead; setting goals and milestones; budgeting time and effort.

- Learning how to learn and adapt in a changing world.
- Working with others.

To measure, reward, and promote all of these competencies, we need:

2) Multiple forms of examination

We want an examination system that will serve a diverse population and recognize many kinds of skills. The traditional forms of examining — usually within a written test on a set occasion — are too narrow and limiting. We envisage three broad classes of examinations that, together, will do the job. They are performance examinations, portfolios, and projects.

> *Performance examinations.* This is the kind of assessment used in the Olympics and in the performing arts. The candidate is given a task to carry out in a limited amount of time, in a special setting. Judges watch the performance and grade various aspects of it, as well as its overall quality. Judges' ratings are then pooled to provide a final grade for the candidate. The essential features of a performance exam are:

- Candidates know the kind of performance that will be called for.
- The performance is an integral one, not a fragmented collection of separate, routine items.
- Students can practice for the performance, with guidance and coaching.

Traditional essay tests are a form of performance exam. But other kinds of "performance" that closely match the requirements of real work and real life can be assessed in this way, too. For instance, practical literacy could be assessed by asking students to put together equipment from written instructions and diagrams.

> *Portfolios.* Judges can also rate work that students have done in school over a period of time, and submitted as part of a portfolio. Portfolio reviews are often used today in the visual arts. They are equally good for any kind of academic skill, such as writing, that can't be assessed very accurately on a single, timed test.

Current experiments in some states show that portfolio assessments can have direct educational value. They can help to focus attention on day-to-day coursework, because any paper or report that a student works on could go into his or her portfolio. They also help to build judgment and an understanding of what constitutes "good work," as students work with their teachers to choose the best of what they've done for the portfolio.

> *Projects.* This form of examination is like the process used for judging student projects submitted to science fairs. The judges evaluate the student's extended work on a task that has real meaning and potential consequences.

Projects in many subject areas can be assigned and assessed. In history or civics, for example, a project might be to investigate and report on a local site

proposed for designation as an historic landmark — or on issues surrounding water allocation in times of drought.

Projects can be done by groups as well as by individuals, giving a chance to assess the student's ability to work with others. They can be undertaken as part of community service programs or work internship assignments. In any case, records would be kept of major steps in the project, and supervisors would sign off on the work and rate it at various stages.

When project assessments are combined with performance exams and portfolio reviews, the result is a very rich evaluation of a student's capabilities.

The third criterion for a good examination system, in my view, is that we must have:

3) A system designed for all, not for an elite

"Excellence" and "equity" are often talked about as if they were separate issues, or even mutually exclusive. Many people believe that we can have education systems which cultivate high standards, or we can systems which are fair and egalitarian, but we can't have both at the same time.

To reach our national goals, however, we must have both. To have a "high standard of living," in every sense of the phrase, we need a nation full of people who are capable of meeting high standards. We can no longer allow a large underclass to exist. The challenge is to design an examination system that will help to promote high standards for all — and enable virtually all to reach them.

• *Cumulative examination credits* are one way of doing this. The system would allow students to assemble exam credits over a period of years, rather than sitting for a single exam (or battery of exams) at the end of each year. Key transitions in school — say, from elementary to middle school, or from middle school to high school — would then be based on a student's having earned specific set of exam credits. The idea is similar to Scouting's merit badge system, where young people work at various paces and various times to gather credentials and move up in rank.

A cumulative system has several benefits. First, it helps to organize and motivate students over a period of years. Instead of preparing for a distant exam whose form and demands can be only dimly imgained by younger children, students can begin early and gradually to collect exam credits.

Second, the cumulative system gives the student multiple opportunities for success, rather than single high-stakes moments of possible failure. Some students may assemble their exam credits more slowly than others — perhaps trying more than one time — while others go faster (and perhaps move on to more advanced, optional work). But all can assemble the credits, and the standards for each credit can remain high.

Instead of having all students move through school at roughly the same pace, but letting some students work to lower standards than others, the cumulative system would keep the standards high for all, but vary the pace at which they're met. This, in turn, avoids the problem of minimum standards becoming functionally the maximum.

• The combined goals of equity and excellence can also be served by *clearly communicating what is expected* of poor and minority students. Again and again, parents of such students are told they're doing well, when in fact teachers are applying a standard to them that is below that for other students of whom more is expected. It is common for poor and minority students to find out much too late that they haven't taken the right courses to get into competitive colleges. The system we have in mind would make it clear from the start what the standards are for everyone, and what it will take to achieve various outcomes after graduating.

The next criterion for a good examination system is that it should have:

4) Shared high standards, but multiple examinations

If a national examination system is instituted, the states and local school districts must still retain ultimate charge of granting diplomas and certifying achievement. This is politically necessary, because education is a function reserved to the states, and local control over education is jealously guarded. It also makes educational sense, because innovation and renewal must permeate the education system, and this is not likely to happen when a single central authority controls exams.

The way to rectify the apparent conflict here is to have a national system of "anchor" examinations, with a procedure for calibrating state or local exams to the national standards. Anchor examinations are exams that directly reflect widely-agreed-upon national standards, so that everyone can see and understand them. Individual states could give the anchor exams directly to their students if they wished. Or, they could develop and use their own exams, but calibrate them to the national anchor by predetermined methods.

In this way, states can pursue their own preferences in curriculum and ways of examining, yet be assured their students are meeting national standards that are benchmarked to the best in the world.

A similar requirement is that we must have:

5) Widespread involvement in developing national standards and exams

A single exam (or battery of exams) that is put into place centrally and imposed from above will not work to change and upgrade teaching practice. It will become — as so many standardized tests are now — something to get around. It will be something to appear to do well on without really digging in to learn and teach, a standard that promotes test gamesmanship, not gamesmanship for life.

To really penetrate and change education, exams need to be developed from the bottom up — with participation of communities in setting the goals and objectives, and with participation of educators in developing and grading the exams. This would allow widespread "ownership" of the process, a key element in getting people to buy into any major change in a system. It would

also build a permanent self-renewing force into the system, by letting school districts and other parties conduct ongoing "R&D" in new methodologies which could then gradually find their into the system. if they have merit.

Finally, for a national examination system to work, it is crucial to:

6) Embed the examinations in broader programs of systemic school reform

Exams are a powerful lever for improving education. But they can't do the job alone. The framework that is used to define the standards for exams must be congruent with the professional preparation of teachers ... with curriculum development at the local and state level ... and with the books, tools, and techniques used to teach. No single element of this system is likely to give us the improvement we want in student performance unless the other elements are also implemented. Thus an effective examination system must go hand-in-hand with an effort to put other components in place at the same time.

Recommendations: What Can Be Done Now

In conclusion, I recommend three action steps to be taken promptly.

> Create a national educational standards framework

The logic that I haved laid out here calls for the process of developing exams to begin with careful definition of detailed goas and objectives for learning. Students are to be actively preparing for these exams, so it is important that what they work on be important — that it be what future citizens of our nation need to know and be able to do. Although the specifics of curriculum and teaching methods will be decisions for states and localities to make, a national educational standards framework is needed to provide the targets at which all would aim their efforts.

The Resource Group that I have been chairing for the National Educational Goals Panel has suggested an approach to creating such a framework. It is a mixed strategy of standard-setting, in which discussion moves back and forth between a national group on the one hand and local communities and states on the other. The national group would have representation from all of the constituencies with an interest in education.

Broad consultation with all segments of our population is the only way to build the commitment of all Americans to the hard job of educational change that lies ahead. Indeed, the standard-setting process itself could become part of the educational renewal the nation seeks, if it is carried out in a broad consultative manner. At the same time, to make sure the standard we set is truly competitive, we will need to benchmark our standards to the best in the world. A process for meeting both goals might work something like this:

First, the national group would collect and make available to the states the curriculum standards or guidelines that have been produced by various national curriculum study groups, along with samples of state curriculum

frameworks and frameworks from several other countries. States would set up procedures for review and comment on these frameworks by groups of educators and citizens, preferably at the level of local schools. The sample frameworks would be used to stimulate and guide discussion. All segments of the population interested in education — parents, teachers and other educators, employers, community advocacy groups, and others — would be encouraged to participate.

Based on these discussions, each state would then compile a report on the preferences of its citizens with respect to educational content and skills. These would be returned to the national group. On the basis of the state reports, the national group would prepare a draft proposal for national knowledge and skill objectives. Representatives from all education constituencies would participate here as they had in local discussions. This draft would be returned to the states, which would arrange for review by local committees. Finally, the national group would revise the draft, thus creating a detailed set of knowledge and skill objectives for the nation. This would be offered to states and communities for adoption. .

The national educational standards would need to be reviewed from time to time, in a similar consultative process, to ensure that educational goals and objectives remain matched to the evolving needs of the country. This means that eventually, a national educational standards board of some kind will be needed. But the initial standard-setting process could begin without a board's being formally established. A commission or national study group, perhaps with a mixture of private and public funding, could lead this process while deliberations on the form and function of a permanent board proceeded.

> Create a system of anchor examinations and procedures for calibration.

Anchor examinations are needed to make the national educational standards real and functional. The anchor exams will exemplify the standards, giving educators and students a clear sense of the goals they are working toward. Anchor examinations cannot, strictly speaking, be created until national standards have been developed and adopted.

However, it is advisable to begin right away to design and evaluate exams of the general forms that are likely to work well in the anchoring system, and to experiment with different methods of calibration. A well-funded development and research program, competitively administered by the Department of Education, could do much to advance efforts toward a national examining system. Such a program should include a mixture of technical studies and exam development. It should include investigations of the role of individual schools and groups of educators in exam development, as well as more traditional university-based or research institute projects.

> Create programs of systemic school reform around the examining system.

To dramatically improve American education we will need new curricula and teaching materials, preservice and inservice training programs, new ways of teaching to new standards, and redesigned systems of responsibility

and accountability for educational practice and outcomes. Without these, new standards and exams are not likely to become anything more than yet another measure of the nation's poor levels of achievement.

These programs are, of course, the responsibilities of state and local education authorities, not of the federal government. However, people developing such new systems will need tools and knowledge for their work. The federal government can play a major role here by helping to fund needed development work and programs of research and evaluation.

The nation is ready to mobilize for improved achievement in our schools. On many fronts, the mobilization already has begun. The actions proposed here will establish the foundation for a powerful examination system that can revitalize teaching and learning well into the next century.

Learning Research and Development Center
and the
National Center on Education and the Economy

THE NEW STANDARDS PROJECT

An Overview

A National Examination System

Our object is to create a national examination system, not a single exam. Development of the system would take place in four steps. First, we plan to involve a lot of people in establishing consensus on a framework for student achievement standards. Then that framework will be used to develop an examination and standards for grading it. At the same time, we plan to develop a technique by which examinations developed by others can be calibrated to the reference exam. With these tasks completed, the conditions will have been met under which a National Examination Board could judge whether any given examination meets the national standard, oversee the process of calibration, and continuously update the exam used as a reference standard. In this way, the nation could have a unified examination system without requiring everyone to use the same exam.

Components of the Examination

The exam would consist of three main components: a performance examination, assessments of student projects (including group projects), and assessments of the contents of a portfolio of student work. The performance examination would take place over a period of hours or days. The work on portfolios and projects would be similar in scope to the tasks set by the Scout merit badge system, permitting the student to accumulate the 'badges' over a period of years. They could work at their own pace and choose for themselves the tasks and projects that they would use to demonstrate their competence against a set of published criteria, just as the scouts do.

Topics and Skills

The examination would focus on the skills of thinking, problem solving and the capacity to apply what one knows to the messy, complex problems found in real life. It would call for real mastery of bodies of knowledge. It would assess not only what one could accomplish working alone, but one's capacity to function effectively as a member of a group. The subject matter would encompass reading, writing, listening and speaking, as well as mathematics, the sciences, history and the social sciences, and work skills. The examinations would put a premium on the capacity to integrate knowledge from many of these disciplines in solving problems.

The Idea of a Mastery Standard

The typical American approach is to use tests as sorting devices. An exam is given once and the scores of those who take it are distributed along a curve. In order to have 'winners,' there must be losers. Our proposal is fundamentally different. We call for an examination which sets a high standard of mastery for all students, and permits them to take the exam as often as they like until they pass it. Some components of the exams might be optional, but those that are not must be passed at least at the criterion level and the criterion level would be the same for everyone.

The Idea of a System that Promotes Student Effort

Today's schooling and testing practices promote the idea that it is native talent or family background that matter, not one's own effort to learn and achieve. Only a few students -- those who know early on that they will compete for selective colleges -- have any reason to study hard in school. With only a few exceptions there is no chance for students to work against a known standard with teachers as their coaches, allies and mentors. We plan to build an examination system in which effort clearly pays off. Students passing a final examination in high school and completing all of their required merit badges will receive a certificate that will signify true accomplishment, not just time in the seat. Intermediate versions of the performance examination will be developed for roughly the end of elementary school and the end of middle school so that younger students, as well, could work towards a clear achievement goal. In such a system, students will be able to see that their own efforts can make a real difference and that school is a place to learn and become competent, not just to be labeled as smarter or slower than others.

Examinations that Students Can Study for and Teachers Can Teach To

For an effort-oriented system to work, it is critical that everyone -- teachers, principals, parents and students -- know just what is expected of them. So we propose an 'open' examination system, one in which the questions, as well as many responses judged acceptable, are released as soon as the exams are over. The secrecy normally associated with exams would be gone. Students would be working toward a clear objective with clear criteria for success.

Clear objectives for students does not mean a 'national curriculum.' Properly used, the exam would be part of a reform strategy, embraced by a growing number of states, that pushes decisions about what to teach and how to teach it down to the professionals in the schools. With the objectives for students clear, those professionals would be free to decide for themselves how to help students reach them, producing much more variation in the curriculum and in teaching methods than we have now.

Embedding the Examination System in a Systemic Strategy for Restructuring the Schools

The governors, through the NGA, have called for the restructuring of American education. In essence, this means a system in which agreement is reached on a detailed set of goals for students, instruments are devised to accurately assess student progress toward those objectives, decisions about how best to assist students to reach those objectives are pushed down to the professionals and parents in the school, and the professionals are held accountable for the results of their efforts. The framework that is used to define the objectives for the students must also guide the professional preparation of teachers, the development of curriculum at the school level and the techniques used to teach. No single element of this system -- including examinations -- is likely to produce the desired improvement in student performance unless the others are also implemented. Our approach to the examination system embeds development in a larger effort to put these other exam components in place at the same time.

Setting a High Standard: Starting with a Volunteer Association of States and Districts

It is crucial, in our view, that, whatever national examination system is established, it reflect international standards of performance. It is very unlikely that that will happen if all fifty states and the territories have to agree on the standard. We are also sensitive to the resistance that the idea is likely to encounter if the effort is perceived as led by the federal government. But it is equally important that the development program be perceived to have the kind of legitimacy that comes from strong participation on the part of key officials from general government and education. For all these reasons, our plan, on which we are now embarked, calls for assembling

a volunteer group of states and school districts to guide the initial development effort. All the members will be required, as a condition of participation, to commit themselves to the general principles just outlined, and, in particular, to set the standard for the examination at a world class level. Senior representatives of these jurisdictions -- governors, chief state school officers and superintendents, for example -- will provide the policy guidance for the entire project, along with others. Our two Centers are now well along in the process of assembling the participants in this 'New Standards' consortium.

How the Standards Get Established

Our project is now assembling standards frameworks and assessment materials from all over the world, as well as the United States. These items are the work of national government agencies, states, disciplinary associations, school districts and others. We are sharing those materials and our evaluations of them with a group of teachers, central office personnel, superintendents, chief state school officers, governors' aides, state and local school board members and others who constitute our temporary working party for the project. Using these frameworks for guidance, and in consultation with a wide variety of organizations and constituencies with interest in education goals and standards, the working party will recommend an initial framework for assessment development to the governing board for the project. The governing board will decide what the assessment standards will be. Thus we have in mind a model that is neither bottom up nor top down, but rather both at once.

Timeline and Funding

The John D. and Catherine T. MacArthur Foundation and the Pew Charitable Trusts have provided $2.5 million to fund the first eighteen months of the project. We expect to have initial pilot tests completed within a year and a half, and to have examinations for the core subjects -- reading, writing, speaking, listening and mathematics -- within 3 years. Within seven years a full examination system can be in place. We estimate that it will take three years beyond that before students can be examined and their performance used as the basis of entry to jobs or further education, because it would be both unfair and illegal to use the exams for such purposes unless the students had been adequately prepared for them.

Project Leadership

Michael Cohen and Lauren Resnick are co-directing the project. Daniel Resnick serves as coordinator of examination development. Marc Tucker serves with Cohen and both Resnicks as a member of the project's management team.

Some Questions -- and Answers

The current tests operate as a formidable barrier to minority and low income students. You are proposing to create a high standard that virtually everyone would have to meet to gain access to college, postsecondary training and good jobs. Won't that be an even higher and more effective barrier for minority and poor students?

Minority and low income students have been very badly served by a system in which they are rarely expected to do well, and frequently rewarded for what is in fact mediocre performance. An examination that simply sorts students by measured achievement will reinforce the inequities that already exist.

Our objective is to make it clear what students need to know and be able to do in order to succeed in school and life and then organize the whole system so that almost all students can achieve that standard. It may take some students longer than others, but the standard will be just as high for them. This is the only kind of system that makes sense if we really believe that all children can learn. For that reason, in our view, it is the only kind of system that offers any real hope to minority and low income students.

In the system we have in mind, the students, their parents and their advocates will be armed for the first time with clear information about what their children have to do to succeed, clear information about how well they are doing against that standard as they progress through school and a clear criterion by which to judge the adequacy of the school curriculum. Students, for the first time, would not bear the whole burden of meeting that standard, because the professionals would be held accountable for student progress. The measures themselves would not be limited to timed performance tests, but would include opportunities for students of many different cultural backgrounds to select their own projects and tasks to demonstrate mastery, making it possible for them to choose performances that play to their strengths. Students who need more time to reach the standard would have the time they need.

You've proposed a ten-year development program. The nation's needs are urgent. We need something in three years or less. Can't this be done faster? And, if it can't, why should we wait?

The ten year implementation plan includes seven years to develop and field test the full assessment system -- all subjects for three different age levels -- and an additional three years for teachers to develop and teach a curriculum that will prepare students for the examinations. Thus we envision a process in which it will take seven years -- not ten -- to develop the assessments themselves; any other plan that entails consequences for individual students that depend on their exam performance will have to allow a period of years, just as ours does, for appropriate curricula to be developed and taught all over the country. A plan that does not do that will be defeated in the courts.

It is also the case that key parts of the proposed examination system -- for example, the performance examination in reading, writing and mathematics -- can be made a high priority for development, and be ready within as little as three years. The rest of the system can be built out from there. We know of no responsible group that has proposed a shorter development time for an exam than three years.

Apart from the development costs, it looks as though the costs to administer the system you propose on an on-going basis will be enormous. The nation can't afford it and shouldn't have to. Can anything be done to get the costs down?

BEST COPY AVAILABLE

42-921 0 - 91 - 2

We have not yet done a detailed assessment of the probable costs of operating the system once it is in place, simply because it is impossible to do so until more design work is done on the system itself. The reason, however, that most people assume that the costs will be very high is because we propose to have human beings -- teachers -- do the grading of student work rather than machines. In most countries, this grading of examinations is done routinely by teachers as a part of their regular duties. We believe that the same thing could happen in the United States. The national examining authority would have to work closely with the teachers' unions to produce that outcome, but we have every reason to believe that this approach could work.

What is to prevent parents and teachers from helping students on their portfolios and projects, in which case you will not be evaluating student performance but the performance of the adults in their lives. Your proposal invites cheating. Isn't that a fatal flaw?

It is quite possible to build checks into the system that could be used to detect cheating. Grades on the performance exam, on which no one could assist the students, could, for example be correlated with grades on portfolios and projects and significant discrepancies could be used as the basis for on-site investigation.

Your system relies on the judgment of the graders. The current system relies on scoring by machines, which is reliable and objective. How can you be sure that your system will be fair and how will you convince the public to have any confidence in it?

Many states now require students to take written and composition tests. These are typically graded by two or more readers. If their grades differ significantly, which happens infrequently, the essay is sent to additional readers. The public appears to accept the validity of this system, which is very similar to the one we propose and to the one used in most other countries.

# # #

For more information, contact:

Learning Research and Development Center
University of Pittsburgh
3939 O'Hara Street, Room 824
Pittsburgh, PA 15260
412/624-7485

National Center on Education and the Economy
Main Office:
39 State Street, Suite 500
Rochester, NY 14614
716/546-7620

Washington, DC Office:
1341 G Street, NW, Suite 1020
Washington, DC 20005
202/783-3668

Senator PELL. Now we look forward to hearing from Gordon Ambach, executive director of the Council of Chief State School Officers.

Mr. Ambach.

Mr. AMBACH. Thank you very kindly, Mr. Chairman. Senator Pell, Senator Kassebaum, Senator Wellstone, it is a privilege to be with you.

And may I say at the outset, Mr. Chairman, of our commendation to you for your leadership in the area of dealing with national examinations. As you pointed out, it was some 24 years ago that you began way ahead of most folks in this country to be thinking about this topic, and you referred specifically to the act taken in 1988. I would like to come back to that in just a moment.

I have a prepared statement which I would like to have in the record. I will summarize the key points and I will make certain that I don't repeat the points of my predecessors with which I agree.

This is an extraordinarily important subject. Our answer to the question of a national test is that the solution is in a nationwide testing system. It is a very comprehensive approach that must be followed. And I speak on behalf of both our council—our board has formally approved what I am about to present—and I also speak from the perspective of 20 years as commissioner and executive deputy commissioner of education in the State of New York, which has this Nation's largest single State testing system, including the only large-scale individual examination system; that is, the Regents Examinations, which you referred to.

I believe there is a lot of experience that we need to look back upon with that system and with what has happened in this country, in order to inform the question that is before you.

The system that should be in place for nationwide testing calls for the creation of a national board, and I have specified what I think ought to be the provisions. This should be jointly appointed by the President and the Congress. This is of such order of magnitude that it needs to be at the very highest levels. There should be a formal nomination process which moves toward appointment.

And the principal function of that board is in establishing the standards, the frameworks from which one develops sets of examinations, both program assessments, which are a sampling of students, in order to determine program effectiveness, such as NAEP does now; and a system of individual student examinations.

The solution that has to be crafted for the United States has to be a unique one, where we are trying to put the right mix together of common standards and a variation of the tests toward which we measure progress of students and systems on those standards.

The program assessment aspect of this national system ought to be based primarily on an expansion of NAEP, and we make an urgent plea that what now is the limitation in NAEP, namely that after 1992 there will be no longer an authority for state-by-state NAEP—and remember it has only been in mathematics and in reading in 1990 and in 1992 that we have had NAEP—that you must move immediately to authorize state-by-state NAEP testing after 1992 so that the planning and the funding can go forward in order to make certain that it is an effective testing system.

We would make one other plea, and that is that you expand this authority, you make it possible on a voluntary basis for those local school districts which wish to participate in NAEP and have a sufficient student population to do so might have that opportunity to do so. One takes a look at the New York City system, for example, which is larger than 30 States' systems in this country, and one wonders why they couldn't participate in NAEP if they wished to do and get that information.

As to the individual examination part, my colleagues have spoken to that. We need a system which has a common standard to be established by the board with a variety of different individual examination approaches.

Let me give you an example, Mr. Chairman, members of the committee, of providing this kind of equivalency. Back in the 1970's in New York, my predecessors established the minimum competency test. Everybody has to take it. And what happened is the roof fell in because, in fact, you had students in private and public schools who were taking SAT's and ACT's and Regents Examinations and other tests which were clearly demonstrating that these students were performing well above a minimum competence. There was no reason to test them.

And so we created equivalencies among things like ACT's and SAT's and Regents Examinations and competency tests so that a student could demonstrate that they met the standard without having to take the single test. That is a concept which is here. I am not arguing using the same mix of tests, but the concept is here, and it's an important one.

Let me provide one other kind of example from the Regents Exam that I think argues for a variety of individual exams and not a single test. It is extremely important, as Lauren has pointed out, that the testing be associated with the curriculum, that it be provided on the basis of the appropriate time for sequences and for mastery so that you don't have everyone sitting on the same day to take the same test across the country. You want to be certain that you are testing when it is that the subject has been mastered and accumulate that record so that we know toward a national standard whether a student has in fact realized that. It can be done. It has been demonstrated in systems like the Regents Examination.

My plea is, then, first, create the national board; second, provide the necessary resources and authorization with respect to NAEP; and third, I believe that it's extremely important for the Congress to begin supporting, perhaps through OERI, the necessary research and development work that will enable us to put together the systems for individual examination that we need. That is a very large task when one is trying to work back and forth between various program or performance assessments and various kinds of short-answer or open-ended questions. How to make some kind of an anchoring of that is critical, and it requires the resources that I believe the Federal Government must put in.

Thank you very kindly.

Senator PELL. Thank you very much.

[The prepared statement of Mr. Ambach follows:]

PREPARED STATEMENT OF MR. AMBACH

Chairman Pell, members of the Senate subcommittee, and staff of the subcommittee, thank you for this opportunity to respond to your invitation to testify on "Question of a National Test." At the outset, Mr. Chairman, I recognize and commend you and your colleagues who have been advocates for national examinations long before serious consideration was being given to these issues for most of the Nation I note especially your responsibility for authoring the provision for national examinations in the Hawkins-Stafford Act of 1988 and commend you and members of the subcommittee now for bringing the issues before the Senate in the manner of this hearing.

Testing is at the center of learning. The characteristics and manner of questioning or inquiry guides the form of education. How tests are constructed shapes what is thought and learned and, therefore, decisions about testing must be carefully made. That is especially true now because of the relation of testing to reform and improvement in American education and the location of education decision-making in the United States. Our American pattern of decentralization of authority for education in States and localities is being weighed off against important national goals and purposes for education. Who establishes standards and develops tests is, in many ways, as significant a question as the content of the test. The right combination of common standards with variations of tests which measure progress toward those standards is central to a national solution.

At the very time we search for successful reforms and improvements in education and debate the relative levels of decision-making at different points in the educational structure, we must also adjust to rapidly changing techniques of assessment. While we have a strong need to do things nationally—to improve opportunities for students nationally and take steps to increase our national competence through education—we must be certain to enable and encourage variation, experimentation and innovation that enables us to create and recreate ever better systems of testing and learning in the future while we are putting new standards and tests in place now.

Our response to a "Question of a National Test" is to recommend "A Nationwide Testing System" which has the following three key elements:

(1) A procedure and national entity for setting national standards for student performance, subject by subject.

(2) A system of both program assessments (through sampling of student performance) at national, State and local levels to determine program effectiveness and a system of individual examinations which might be nationwide, multi-state, State or local in administration which measure individual student progress on the national standards.

(3) A reporting system which enables students, parents and responsible education officials anywhere in the United States to be able to relate any one student's achievement to the national standards and the performance of other students in the community, State, Nation, and even the world.

The first element is to establish a procedure and an entity for setting national standards for the nationwide testing system. The United States currently has no national entity to establish national student performance standards. Such an entity, or board, must be carefully designed and established through an act of the U.S. Congress and the President. The board should be comprised of distinguished persons appointed in equal numbers by the Congress and the President. Appointments to the board should be based on a thorough nomination process which assures board members will be well-qualified for their responsibility. The board's responsibility should be to establish "frameworks" of student performance goals and objectives, or standards, upon which both program assessments and individual student examinations are based. The process of setting such frameworks must involve key education authorities at State and local levels.

Although it might seem establishment of such national frameworks is foreign to American education practice, the fact is that such a process is in place for the National Assessment of Educational Progress (NAEP). The process has already been used for the subjects of mathematics, reading and the sciences under the direction of the Council of Chief State School Officers as a part of the development of NAEP examinations. A similar process could be used by the new board.

A second important responsibility of the board should be review of various proposed program assessments and individual examinations in order to determine whether they effectively measure the standards established by the board with reliability and validity. The board would maintain quality control of various tests to assure rigorous measurement of both what students know and can do. The board would exercise judgment on effectiveness of a proposed test for its intended purpose and the design for reporting results on the test.

At this time, Mr. Chairman, we are not presenting a specific bill for the creation of the board, but we would be pleased to assist you and members of the subcommittee and the staff in the development of such a bill.

Program assessment components of the nationwide testing system might be at the national, State and/or local levels. NAEP provides the basis for the major program assessment component. For nearly a quarter century, NAEP has been providing periodic testing of samples of students across the Nation. It provides overall trendlines for student achievement.

In 1988 Congress authorized the use of NAEP on a state-by-state basis. In 1990 the first use of NAEP on a state-by-state basis—mathematics at the eighth grade level—was implemented. You have authorized state-by-state NAEP in mathematics at two grade levels and reading at one grade level in 1992. Authority for state-by-state NAEP, however, stops at that point.

We urge you take action as rapidly as possible to authorize the continuation of NAEP on both a national and state-by-state basis in five major subject areas mathematics, science, reading, writing and history/geography—each to be tested every 2 years. We urge, furthermore, that you authorize voluntary participation in NAEP at a school district-wide level for those districts of sufficient see for appropriate NAEP sampling. We have a marked up bill which would accomplish the amendments recommended above.

Within the overall nationwide testing system, NAEP tests would be constructed to measure the subject objectives and standards established through the new board frameworks described under point (3) above.

A nationwide individual examination system should be established within the nationwide testing system. This system could have several different forms of examinations. These might be used on a nationwide basis or by clusters of States or districts, or individual states or school districts. The determination of their use would be made by state or local education authorities. The examinations would be based on standards established by the board under item (3) above. The types and forms of various examinations would differ, but, as noted above, to be part of the nationwide system they would have to be judged appropriate to measure the national standards by the board. This system would enable creation of innovative forms of testing, including performance assessment, and enable a variety of approaches by states and localities in establishing individual examinations in an efficient and cost-effective manner. These examinations must be closely associated with the curriculum; used in a variety of patterns at different grade levels according to those points study at which subject mastery is completed; and be used as "high stakes" tests related to credit or credentials for individual students.

States and localities have a variety of existing individual examination systems, the most comprehensive now being the New York State Regents examinations. Different systems, such as that in New York, could be incorporated into the nationwide individual examination systems.

Of special note, Mr. Chairman, is the potential incorporation of the voluntary national examination authorized in the Hawkins-Stafford Act of 1988 into this system. This examination has never been implemented. It could be one part of the nationwide individual examination system, its content guided by the standards of a new national board.

We recommend you take action to support the research and development needed to establish a nationwide individual examination system.

Common and consistent reporting of results from the different program assessments and individual examinations is of central importance in the nationwide testing system. There is now extensive testing at all grade levels for American students. The information from the tests, however, generally cannot be related so that results from one school or school district may be compared with another district within the State, or outside of the State, or outside of the Nation. One purpose of the nationwide testing system is to create the means for relating results throughout the system without the necessity of requiring all students at 'all grade levels in all subjects to be tested on the same tests. Reporting systems must be established nationwide with information on student results related to other education indicators. This enables better understanding of the causes of student success or failure and helps results of testing lead to program and student improvement.

Testing systems carry significant costs in student instructional time, teacher time, and in the cost of creation, administration scoring and reporting of test results. Careful estimates must be made on the trade offs between program assessment sampling versus individual examinations and "every student testing." Careful estimates of cost must be made over this decade so that an efficient mix of program assess-

ments and individual examinations is created and coordinated so as to limit costs locally, state-by-state, and Federal.

A nationwide testing system is essential. Information about the Nation's education status is certainly as important as information about the Nation's health, its agricultural condition, and the condition of labor and employment. In each of these other areas, the Federal Government is now spending approximately six times as much for the collection of information about performance and system indicators as is true in education. The commitment to a nationwide testing system must be accompanied by a commitment to a Federal budget which makes certain the testing system is of as high quality as we expect student performance to be.

Mr. Chairman and members of the subcommittee, thank you for the opportunity to testify. I will be pleased to respond to any questions.

Senator PELL. Mr. Jack Anderson, superintendent of schools, East Ramapo Central School District, Spring Valley, NY.

Mr. ANDERSON. Good morning, Mr. Chairman, members of the committee. On behalf of the American Association of School Administrators, I want to thank you for the opportunity to testify today on what is certainly a most important issue for education.

I am superintendent of the East Ramapo Central School District in Spring Valley, NY, and my district has an enrollment of 18,000 public and nonpublic school students, with a professional staff of some 1,000 teachers and administrators.

My remarks will reflect both the position of AASA and my own, which has been formed over some 30 years in education. The American Association of School Administrators is a professional organization comprised of over 19,000 local superintendents, local administrators, and professors of higher education in the education field.

While we have presented written testimony, I would like to speak to a couple of issues this morning.

First, can a national test be a major force in improving education? The short answer is: No, testing cannot be a major force in producing quality education. The tests themselves are often suspect. Tests may not test what is being taught in a particular school district, as our some 15,000 school districts across this country do not necessarily introduce the same concepts or follow the same curriculum scope and sequence through the various grade levels.

It is well founded that no single test can respond to the various learning styles found in any large population of students. Since there is no national consensus about what students should know and when they should know it, it follows that wide-scale tests do not measure a definitive body of knowledge that defines the national interest in education.

Tests are all too often discriminatory in that they hold students responsible for conditions that they neither created nor have the ability to control. In general, youth who live in poverty, Afro-Americans, and Hispanics score at lower levels than economically advantaged white students. At the same time, Asian students score higher than whites, Afro-Americans, and Hispanics. It would be all to simplistic to suggest that test scores based on school experiences explain these diverse findings.

Testing, for the most part, occurs outside the daily learning environment. Frequently, results are not important to students, and quite often the presentation of results is often separated in time and format from the actual test and classroom activities.

But more importantly, the notion that quality can somehow be achieved through inspection does not and will not work in education any more than it has worked in industry.

Second, can test scores be used as a prod to force educators to do a better job? Again, the answer is no. There is in some sectors of our society today a belief that performance criticism is a powerful tool in behavioral modification. Experience, however, shows that this is simply not the case.

A national test used to compare educational programs at the local, State, and regional levels will generate the same lack of results and, in fact, respect as the Department of Education's so-called wall charts.

Third, can information about what students know and be able to do be useful in both the classroom and in the political arena? The answer is yes. Though there are a few models in place, a number of States are now developing information systems about student learning that can serve both the need for information on what has been learned and the need for political leaders to understand the impact of educational policies.

Finally, an AASA proposal to improve both the information important to local educators and information desired by national policymakers. Rather than thinking about testing or a test as a cure for the Nation's educational problems, real or imagined, we need a national consensus on what children are expected to know and be able to do when they graduate from high school. We do not need an exhaustive list of outcomes, but rather, broad standards. These standards should most definitely not be a curriculum and most emphatically should not be a test. Rather, they should describe what students should know and be able to do. Standards, not standardization, guarantees, in my opinion, student achievement.

AASA strongly suggests that Congress form a commission to develop three major tasks:

(1) Develop criteria to consider and judge testing, assessment, and accountability models;

(2) Develop a national consensus regarding the level of student achievement for high school graduates; and

(3) Identify indicators of child development that will provide a barometer of the status of children and youth.

On a final note, certainly significant to the overall improvement of education, it is urged that Congress establish a Marshall Plan for children, directed toward identifying and attacking the at-risk issues facing hundreds of thousands of children in this Nation.

Thank you, Mr. Chairman.

Senator PELL. Thank you very much indeed, Mr. Anderson.

[The prepared statement of Mr. Anderson follows:]

PREPARED STATEMENT OF JACK R. ANDERSON

Mr. Chairman, on behalf of the American Association of School Administrators, I would like to thank you for the opportunity to testify today on what is certainly a most important issue for education. I am Superintendent of the East Ramapo Central School District in Spring Valley, New York. My district has an enrollment of 18,000 public and nonpublic school students, with a professional staff of 950 teachers and administrators.

My remarks will reflect both the position of AASA and my own, which has been formed over some 30 years in education. The American Association of School Administrators is a professional organization comprised of over 19,000 local superintendents, other local administrators and professors of education administration.

Currently, there are a number of organizations and prominent individuals who support the position that either more tests or, in this case, a national test is necessary to the improvement of education and in order to hold local educators accountable for the education of students. AASA would like to address three major issues related to these positions and, particularly, to the important matter of testing, assessment and accountability in our public schools.

The first issue that arises is whether or not a national test can improve education. Secondly, an important question related to accountability is, simply stated, can education be improved by publishing comparative test scores that (as we know from experience) causes public discomfort for students, teachers, administrators and parents in those schools and districts that

BEST COPY AVAILABLE

produce lower test scores. And thirdly, can information regarding what students know about the same thing, at the same time, and in the same way be useful to improving their education and the political process.

These questions are not all inclusive regarding assessment, testing and accountability, but they seem to be extremely central and most relevant to the matter of a national test.

I. CAN A NATIONAL TEST BE A MAJOR FORCE IN IMPROVING EDUCATION?

The short answer is no; testing cannot be a major force in producing quality education, for at least three reasons. First, the tests themselves are often suspect. Tests may not test what is being taught in a particular school district, as all districts do not introduce the same concepts or follow the same curriculum scope and sequence through the various grade levels. Secondly, we know from a significant body of educational research that multiple intelligences show clearly that how questions are asked has a great deal to do with how well students perform. It is well founded that no single test can respond to the various learning styles found in any large population of students. Since there is no national consensus about what students should know and when they should know it, it follows that wide-scale tests do not measure a definitive body of knowledge that defines the national interest in education. While educators recognize that different learning styles exist and are important in test taking, the tests do not recognize these differences. Yet, the measurement of educational program success or failure is all too often

based on narrow testing procedures and test items. We must recognize that a major responsibility of our schools is to meet the many and varied learning styles of our student population, and in doing so apply appropriate assessment techniques to those styles in order to measure success.

We should not and, in fact, cannot fail to answer important questions as to the most appropriate way to assess and motivate gifted students, the handicapped, minorities and multicultural student populations that have special assessment needs. National testing would, in all probability, be another layer of "test and compare" activities that make no contribution to the improvement of education.

Tests are all too often discriminatory in that they hold students responsible for conditions that they neither created nor have the ability to control. In general, youth who live in poverty, Afro-Americans and Hispanics score at lower levels than economically advantaged White students. At the same time, Asian students score higher than Whites, Afro-Americans and Hispanics. It would be simplistic to suggest that test scores based on school experience explain such diverse findings. Cultural effects and socioeconomic conditions are very controversial, yet critical in the consideration of teaching/learning and the measurement of academic accomplishments.

Suffice it to say that students who live in poverty, diminished family environment, and those who experience limited personal accomplishments in their total environment will not attain the same level of educational accomplishments. We know that

children represent the largest at-risk population cohort in our nation, and these children living in poverty come from poverty-ridden homes. Unfortunately, schools do not have the necessary human and fiscal resources to do what most educators know would lead to greater success for this student population. Therefore, we knowingly allow students at risk to fall by the wayside of education. Simply stated, we know that all children can learn, but we also know that all children are not being given the opportunity to learn. A major myth of education is that the schools can correct deficiencies in education by simply testing in order to point out what deficiencies exist. Such an approach is clearly doomed to failure. We must recognize that those societal shortcomings that impact negatively on student achievement and for which schools are not responsible cannot be corrected by changing and assessing the educational process.

We must further understand that agencies concerned with social services, health services, the homeless, child day care and parent training, among others, must be brought together in a concerted and united coalition with education if we are to make a significant and positive impact on the improvement of education for all children in this nation. We should not be testing until we know what we really want for our children and have in place strategies to accomplish those goals.

Third, testing, for the most part, occurs outside of the daily learning environment. Frequently, results are not important to students and, quite often, the presentation of results is often separated in time and format from the actual test and

will be used to compare educational programs at the local, state and regional levels and, I must say, generate the same lack of results.

This new layer of national testing will overlay on a host of tests already being given at the state and local levels. Many states, such as New York, publish a Comprehensive Assessment Report on academic achievement, district-by-district and school-by-school each year. This report has turned into a media event with no positive implications for the improvement of education. A general public that is not sophisticated (understandably so) in testing draws false and damaging conclusions regarding these results and their schools. Knowing that this is the game that will be played with national tests, educators, through their experience at the state and local levels, will either choose to teach to the test or ignore it as a meaningless waste of time and effort. Both attitudes tend to develop and perpetuate mediocrity of curriculum and teaching with a significant loss of creativity.

It should be noted that currently there are school districts in the State of New York already utilizing four weeks of an already too short school year in order to administer a vast array of national, state and local tests!

Finally, thinking that embarrassment through publishing individual test results can produce improved education assumes that teachers and administrators are sitting with their motors idling when it comes to professional practices, and that a national test will cause them to get their act in gear. To the contrary -- there is overwhelming evidence that the vast majority

will be used to compare educational programs at the local, state and regional levels and, I must say, generate the same lack of results.

This new layer of national testing will overlay on a host of tests already being given at the state and local levels. Many states, such as New York, publish a Comprehensive Assessment Report on academic achievement, district-by-district and school-by-school each year. This report has turned into a media event with no positive implications for the improvement of education. A general public that is not sophisticated (understandably so) in testing draws false and damaging conclusions regarding these results and their schools. Knowing that this is the game that will be played with national tests, educators, through their experience at the state and local levels, will either choose to teach to the test or ignore it as a meaningless waste of time and effort. Both attitudes tend to develop and perpetuate mediocrity of curriculum and teaching with a significant loss of creativity.

It should be noted that currently there are school districts in the State of New York already utilizing four weeks of an already too short school year in order to administer a vast array of national, state and local tests!

Finally, thinking that embarrassment through publishing individual test results can produce improved education assumes that teachers and administrators are sitting with their motors idling when it comes to professional practices, and that a national test will cause them to get their act in gear. To the contrary -- there is overwhelming evidence that the vast majority

classroom activities. But more importantly, the notion that quality can somehow be achieved through inspection does not, and will not, work in education any more than it has worked in industry.

Quality cannot be inspected into a product; rather, quality can only be infused through an improved process recognizing the vast array of factors related to learning. Industries that have tried to produce quality through tough inspections, while ignoring the human and social components of the assembly line, are not successful. Education has, and will continue to have, the same results if we use the same approach of measuring outcomes, while ignoring process needs and components.

II. CAN TEST SCORES BE USED AS A PROD TO "FORCE" EDUCATORS TO DO A BETTER JOB?

Again, the short answer is no. There are in some sectors of our society today a belief that performance criticism is a powerful tool in behavior modification. Teachers and administrators whose students do not score well can be professionally embarrassed; however, in the absence of clear definitions regarding the causes of failure, improvements cannot, and will not, occur. Understanding student needs and how to address them is better done by teachers who work with students daily than by panels of experts operating in isolation from the grass-roots process in our state capitols and in Washington, D.C.

Despite predictable protests to the contrary, a national test (like the so-called Department of Education "wall charts")

of teachers and principals perform to a very high level day in and day out despite, in some cases, overwhelming odds against their success and the success of their students. Standardized tests simply do not measure teacher or administrative competency.

It is important to note that at a time when many states and the Congress of the United States is looking to implement initiatives that would assess the success or failures of education, we have disastrous reductions in monetary support for education at the state and local levels.

The State of New York is an excellent example. While the State Education Department is developing new educational initiatives, the Governor has proposed an $891 million reduction in aid to public schools in the 1991-92 fiscal budget. This, coupled with a $191 million reduction during the 90/91 school year, equates to a two-year loss of support for education in excess of one billion dollars. Many districts in New York are eliminating all but the most core educational programs, and literally thousands of staff reductions are occurring as class sizes drastically increase and programs for at-risk students are eliminated. The same condition exists today in many of our states.

III. CAN INFORMATION ABOUT WHAT STUDENTS KNOW AND ARE ABLE TO DO BE USEFUL IN BOTH THE CLASSROOM AND POLITICAL ARENAS?

The short answer is probably yes. Though there are few such models in place, a number of states are now developing information systems about student learning that can serve both the need for daily information on what has been learned and the need of

gains can be described in terms of locally developed outcomes utilizing statewide or commercially available tests that relate directly to the scope and sequence of local curriculum and goals of the district. Disaggregated data relating to the student population data base is also critical in the assessment of student achievement.

IV. AASA PROPOSAL TO IMPROVE BOTH THE INFORMATION IMPORTANT TO LOCAL EDUCATORS AND INFORMATION DESIRED BY NATIONAL POLICYMAKERS.

Rather than thinking about testing or a test as a cure for the nation's educational problem, real or imagined, we need a national consensus on what students are expected to know and be able to do when they graduate from high school. The National Council of Teachers of Mathematics has developed a splendid set of "standards" for mathematics that can serve as an excellent model. The science, English, social studies and reading teachers have done, or presently have underway, similar work. We should capitalize on these outstanding efforts, as well as the work of NAEP and various academic commissions to develop a national consensus on standards for education.

We do not need an exhaustive list of desired outcomes, but rather the type of broad standards identified as a "National Standards Framework" by Lauren Resnick, recently in testimony to the Romer Panel on Implementation of the National Educational Goals. These standards are most definitely not a curriculum and,

most emphatically, not a test. Rather, they describe what students should know and be able to do. States and school districts can then determine how best to achieve these standards.

If employers could have accurate descriptions of what students know and can do, and if all students have achieved these standards, there would be little use to compare student tests that rate and rank students. AASA sees little reason to sort and separate students, because such practices are inconsistent with the notion that all students can learn. Testing to sort and separate students is also inconsistent with the philosophy of not holding students responsible for factors they did not create and do not control.

Therefore, AASA strongly suggests that Congress form a commission to accomplish three major tasks:

(1) Develop criteria to consider and judge testing, assessment and accountability proposals. These criteria would provide a consistent model for state and federal policymakers and local school boards as they develop new methods of accountability.

(2) Develop a national consensus regarding the level of student achievement for high school graduates. We suggest using the standards of the various curriculum associations mentioned previously as the jumping-off point in developing the consensus on a standards framework. Again, the standards are emphatically not a curriculum or a test. We should emphasize that reform starts with knowing where you want to go, and success depends on focusing on teaching strategies that enable students to reach these standards.

BEST COPY AVAILABLE

(3) Identify indicators of child development that will provide a barometer of the status of children and youth. A regular report on the status of children and youth must accomplish any public review of learning; otherwise we can never begin to understand the impact of social, economic and cultural forces on children's lives as it relates to their educational development.

On a final note, certainly significant to the overall improvement of education, it is urged that Congress take the lead in providing the impetus for an aggressive coalition effort of all those agencies and organizations impacting on the education and social fabric of our nation's children -- an effort directed toward improving on the chances of success for all our children. A National Child Recovery Act directed toward recognizing and attacking the at-risk issues facing hundreds of thousands of children in this nation is of the highest priority.

On behalf of AASA, I thank you for the opportunity to express these views regarding the important area of educational assessment and accountability.

Senator PELL. I have just one question. That is, should the tests be mandatory? I think we all agree it would be helpful to have a test. Should it be mandatory, or should it be optional? I would just ask first, Mr. Kolberg.

Mr. KOLBERG. Well, I would repeat, Mr. Chairman, what I said earlier. I think we need a set of national standards, and I think the standards need to be, as I think my colleagues have said, not standardized in the usual sense of terribly prescriptive, but as Mr. Anderson has just said, we need to know what high school students are expected to do and perform and the knowledge that they need to have; the competencies, in other words. And that comes about by setting standards.

Business needs to become much more involved in competency standard-setting. Certainly the schools need to do that, and we need to have that kind of a system across the United States. And, I think, in one way or another, all four of us have been saying that.

Senator PELL. But, in brief, it should be mandatory?

Mr. KOLBERG. I think they should be mandatory.

Senator PELL. All right.

Mr. KOLBERG. I think this society should expect—and a commission that Lauren and I happen to serve on, the Commission on Skills of the American Workforce, calls a certificate of initial mastery—we ought to, by the age of 16, specify what the performance standards, competency standards are and expect and work with and do whatever needs to be done to help all Americans reach those standards.

Senator PELL. Thank you.

Dr. Resnick.

Ms. RESNICK. Well, I agree that students ought to be required at some key points, two or at most three, in their careers to pass a real mastery criterion examination. I think that a single national test for that purpose would be counterproductive. I think it will become the kind of standardization that Mr. Anderson referred to and will not be a vehicle for energizing and recruiting the attention of not only teachers but parents, communities, to this problem that we need.

The notion of multiple examinations calibrated to a single standard is a system that will work for this country. I strongly urge that we go that route rather than the route of a single national exam or test.

Senator PELL. Thank you very much.

Mr. Ambach.

Mr. AMBACH. Mr. Chairman, the requirement should be that there is a system for reporting on each child or student with respect to that student's progress toward the national standard. That is a very different requirement than requiring a single test of everyone. But it is a requirement that there would be a report in knowledge of how each student is performing on that standard. The way in which one determines that report or that information would be from a variety of different exams.

So the requirement would be on the reporting, the demonstration of progress, rather than on a single test that would have to be taken.

Senator PELL. Thank you very much.

Mr. Anderson.

Mr. ANDERSON. Mr. Chairman, as a local superintendent, I am very concerned that whatever the Congress decides to do, that we are able to do in a productive way, in a meaningful way to help children to better be prepared.

My concern at this time is that we really need to identify those standards. We then need to spend a lot of time determining how we are going to test not only in the cognitive skill area but also in the affective domain in terms of the thinking process which is going to be critical to education in the 21st Century, in my opinion.

So, we feel the test at this time is premature to a lot of important work.

Senator PELL. Thank you very much indeed.

Senator Kassebaum.

OPENING STATEMENT OF SENATOR KASSEBAUM

Senator KASSEBAUM. Thank you, Mr. Chairman.

First, let me say I do have an opening statement which I would like to make part of the record.

Second, I guess I am glad I grew up in an age where we didn't have to worry so much about all these tests. I am not sure I would have made it through.

I think that, in general terms, part of this growing interest in national testing, along with many other things that we are debating regarding education, stems from our loss of confidence in our education system. So we are sort of reaching out trying to figure out what might work, and I think sometimes we miss the forest for the trees.

But it's very interesting to hear these observations. I couldn't agree more that we need to—and I have spoken to this strongly at home—expect more from out students today. I think each school district can establish higher standards for graduation from high school. This doesn't need to come from Congress. It doesn't necessarily, I would suggest—Mr. Anderson, you know this far better than I as a superintendent—mean more hours in the classroom or more courses. It's just expectations being raised.

[The prepared statement of Senator Kassebaum follows:]

PREPARED STATEMENT OF SENATOR KASSEBAUM

It is a pleasure to welcome today's witnesses as the education subcommittee examines the question of national testing.

This hearing is a timely one. Certainly, the recent momentum on behalf of some form of national testing could scarcely have been imagined even a few years ago.

The growing interest in this subject reflects a recognition that we have a stake as a Nation in producing well-educated citizens and workers. Unfortunately, it reflects as well a loss of confidence in our education system—a sense that the high school diploma, once a demonstration of a basic level of proficiency, no longer adequately serves that purpose.

Believing, as I do, that we need to increase our standards and expectations for students, I have some sympathy for the concerns expressed by national test proponents. At the same time, I am one

who has felt that one of the real strengths of our education system is the fact that its roots are firmly planted in local soil.

Our public elementary and secondary school system stands as one of the most truly democratic institutions in our society. Key decision makers are readily accessible to the local community and to parents—whose role in promoting a respect for education and an insistence on achievement is irreplaceable.

Consequently, I approach the question of national testing with a certain amount of misgiving—particularly in terms of reconciling the concept of some type of national standard with that of the traditional (and important) local focus of education.

It is important, I believe, to address several key questions:

—Would, in fact, a national test lead to a national curriculum?

—Would the results of a national test be meaningful to the classroom teacher, the individual student, or his/her parents? Or would it, rather, be one more distraction from regular classroom work?

—Would a national test actually result in raising our standards and, if so, how?

—Given the expense of national testing, is this the best use of resources spent for school improvement efforts?

—Who decides what will be tested—or, more to the point, who decides what is important for our children to know?

The witnesses appearing today hold diverse viewpoints not only about the desirability of national testing but also about how any such test should be structured.

I look forward to a lively exchange as well as an opportunity to learn.

I would like to ask some questions of each member of the panel because I do think that it is a fascinating subject.

Mr. Kolberg, I will start with you. How do you believe we should go about establishing the consensus you mentioned?

And, Dr. Resnick, I think you spoke to this in many ways, too, in terms of the technicalities of getting this all put together. That in and of itself could take a lot of time and money and effort. Is it going to drain a certain amount of vitality from where I wish our focus might be on this?

And then second, let me ask you, do you think a national test would lead to a national curriculum?

Mr. KOLBERG. I am afraid too often we emphasize the pessimistic rather than the optimistic. This panel, I think, shows the optimism that all of us in our various ways have come a great deal of way over the last three or 4 years in understanding the issues in educational reform, and you find us coming from very different places in this whole society but saying a lot of the very same things. So I think we have made a lot of progress.

But I would repeat what I said in my statement. I hope that the new Secretary of Education, when he is confirmed by this body, will be the kind of proactive leader that I think you need. I think a Nation needs a proactive leader and you need a cabinet officer who can speak on behalf to the education president every day and lead us through a variety of experiences together, whether we set up boards as some of us suggested, whether we strengthen pieces of the Education Department. I think it's that kind of leadership that

is needed to build the national consensus around what we have been talking about here today.

Senator KASSEBAUM. Well, let me ask all of the panelists, do you think a national testing would lead to a national curriculum?

Ms. RESNICK. I have, and I think my colleagues also, are not talking about a national test. We are talking about national standards, and we are talking about, or I am anyway, a process of standard-setting that wouldn't be draining energy, it would be part of the process of mobilizing the country for really improved educational activity.

I have been chairing a resource group for the National Educational Goals Panel, made up of people from very different entering points of view, whom I did not get to a point even though I was chairing it, and we have suggested to the panel an approach to this standard-setting process that perhaps I can share with you.

It is designed to be a mixture of top-down and bottom-up activity with broad consultation with people. We imagine a national group of some kind, perhaps a board but it could just be a commission or an appropriately vested study group, that would collect and make available to each State the curriculum standards and guidelines that have been produced by a number of national curriculum study groups along with samples of State curricula and frameworks from other countries, because this is, after all, partly an international competitive engagement that we are in.

States would then set up procedures for review and comment on these frameworks by groups of educators and citizens, preferably at the level of local schools.

The sample frameworks will be used to stimulate and guide that discussion, but it would not be imposed on anyone.

As a result of this, the States would be able to compile reports of what their citizens want for their students, and that could come back to a national commission that could draft a proposal for knowledge and skill objectives that would be much more detailed and usable than the current national goals that we have.

That draft can then go back down through the process and back up. The details of this need to be carefully designed. What our group did was try to outline an approach that we believe could mobilize attention and commitment to the educational problem that we have.

This is not a process of setting a national curriculum. It is a process of getting the whole country involved in seriously thinking about what the educational objectives ought to be.

Senator KASSEBAUM. Perhaps this is a little simplistic, but wouldn't we all recognize a really top-flight student, without having any exam results before us? Mr. Anderson.

Mr. ANDERSON. You are absolutely correct. There is no question that I can predict for the Congress or anyone else the student cohort population in my school district that will succeed at the high level, that will be at the middle level, and those that will fail. And we will tell those who fail, once again, you are not succeeding in our society. And we do too much of that today.

In our school district, Senator, we have over 4 weeks of testing already per year in an already too-short 180-day school year. Now, when you add the teacher time that is spent preparing students for

those 4 weeks of testing, look at the amount of time that we are taking out of the instructional day in order to assess.

So we have to be careful about overlaying these tests. We have to have some coordination if we come from the national level with the different States. We have to have an understanding in terms of how much time should be spent on testing.

To your question in terms of a national curriculum, I think it follows that if we have national standards—and I support national standards—those standards will equate then to a taxonomy of skills, if you will, to be taught in the classroom and de facto almost you will be impinging in a positive way on a curriculum on a national scale.

But to develop a national syllabus, I would be opposed to that.

Mr. AMBACH. Senator, excuse me, if I may?

Senator KASSEBAUM. You may.

Mr. AMBACH. Could I respond to the coupling of questions that you first asked to Mr. Kolberg. One was about the consensus process and the other is about a national curriculum.

One might think that the matter of developing a framework or standards for national testing in this country was completely antithetical to our system. The fact of the matter is that we have been doing it for the last 25 years, in the NAEP program. And within the last three or 4 years that process has become much more sophisticated by way of the development of what are called frameworks, which means the set of objectives and goals or standards from which the test is drawn.

Now, under NAEP, and that is program assessment, not individual exam, that process has been carried out by our council. It is a consensus process which has involved representatives of all different levels of the educational community, of representatives of the public, and we have through that process developed the objectives for the testing in NAEP.

So we have already examples of where it is possible to develop a kind of consensus, if you will, for the country as to what the standards should be. That process needs to be made more sophisticated. It needs to be, in my judgment, coupled in with what I proposed for a national board. But I think you would find that we are lot closer to having the capacity and the acceptability of developing this kind of set of standards than many might have thought.

Now, that is not, in my judgment, the creation of a national curriculum, nor is it the creation of a syllabus, nor should it be. And I believe that that is the pivotal issue before you by way of how you would exercise Federal responsibility in establishing any kind of a board or mechanism and how you design it, so that we are certain it is a standard-setting process but it is not a process which in fact sets the curriculum or a single test to try to test it.

Senator KASSEBAUM. I have used up my time, but let me just ask because all four of you are saying the same thing regarding that, I believe. Don't you believe, looking at past tests, that eventually a teacher starts to teach those things that are going to be? So we are beginning, then, I would think, to focus on some kind of national consensus.

Now, maybe it's something that is important to focus on, and that does become the curriculum, to a certain extent. For example,

there are certain facets of history that everybody should know, and maybe that isn't all that bad. I do think the evolution of it could become a national curriculum.

Mr. KOLBERG. Well, Senator, it seems to me as a noneducator, it seems to me that you could start with a national curriculum and more then to some kind of national assessments as well as start, as we are talking about today, move to develop a set of assessments and then ultimately build a curriculum around it.

The British have recently chosen to go at it, by my layman's view, more directly and say we want a national curriculum. And ultimately and inexorably, out of that national curriculum will develop the kinds of national assessments that we need.

I use that as an illustration because it seems to me that they have taken a terribly bold step, for that society, and we should not stand back from thinking bold thoughts if we are serious about meeting our national goals.

Senator KASSEBAUM. Thank you.

Thank you, Mr. Chairman.

Senator PELL. Thank you very much.

Senator Bingaman.

Senator BINGAMAN. Thank you very much, Mr. Chairman. I apologize for not being here for all the testimony this morning, but let me just ask one question.

Mr. Ambach, let me ask you, you and I have talked about this issue many times and I have always appreciated your point of view and expertise. I notice on page 3 of your testimony that you talk about the first element being the need to establish an entity for setting national standards. That would be a national entity to establish national student performance standards.

Are you clear, is this an entity similar to what we are contemplating with the goal-setting mechanism that I have proposed in Senate bill 2 or that is included as part of that legislation? Or is this something different?

Mr. AMBACH. Senator, this is something different. Again, you and I have had a number of exchanges. It may seem as though I have been a bit inconsistent, because, as you well know, I have argued very strongly that with respect to the panel which has been established by the President and the governors on monitoring the goals and with respect to anything like that, that there should be limitations in that panel and that its limitation should be strictly a limit of monitoring progress on the goals.

In my judgment, we are talking about something different here. In my judgment, we are talking here about the establishment of sets of standards subject by subject of what we would expect American children to know and to be able to do. And I think that is a function which is beyond that more limited function of actually tracing progress on the goals. In fact, it is closer, sir, to where you were originally with the panel recommendation, which had that broader scope and was not limited to monitoring.

Senator BINGAMAN. But you are recommending essentially that somebody takes the basic core subjects that we expect our students to learn about and identify very detailed, specific standards of performance we expect them to achieve in each of those subjects.

Mr. AMBACH. Sufficiently detailed so that you can then use them to design either program assessment instruments or individual examinations.

Senator BINGAMAN. And let me just be clear in my own mind. How do you propose this would best be organized? Would it be under the Department of Education? Would it be independent of that? Would it be a combination? How do you see that?

Mr. AMBACH. My recommendation, Senator, is that this be an independent of the department, this in fact be a body which would be appointed in even parts by the President and the Congress, some by the House and some by the Senate; and that, in effect, it would be different, for example, from the NAGB board, which is currently now appointed by the Secretary of Education. That is not a criticism of the NAGB board or a criticism of the Secretary. There is a different function.

I believe we are now dealing with much higher stakes than we ever have before with respect to establishing the appropriate governmental board or entity that needs to deal with this. And it needs to be that it can have the kind of independence which I believe could be held if there is that jointly appointed arrangement.

I am considering here, although I haven't given you all the details, of terms which would be 5 to 7 years. You would want them to bridge over. I think there would be a necessity of having very careful provisions with respect to assuring that this is not a partisan body. I think there needs to be a very careful process with respect to the manner of reviews or nominations for appointments.

We are talking here about, in Bill Kolberg's words, moon shots or major efforts of serious development for education in this country, about something that we have never had before.

Senator BINGAMAN. That is all I have, Mr. Chairman.

Thank you very much.

Senator PELL. Thank you very much.

Senator Wellstone.

Senator WELLSTONE. Thank you very much.

I have, one more or less statement. It kind of follows on the heels of Senator Kassebaum's questioning, and then one question which I am very interested in your response to.

The statement, question, or whatever you want to call it, is that I at first I thought we were talking about a single mandatory national test. I have heard something a little different. I am glad to hear something a little different. I have heard language about a system of reporting, and then I think it was Dr. Resnick who talked about a process whereby we begin to formulate national educational objectives.

For my own part, I guess I, at some point in time before making a final decision, want to be clear exactly what we are talking about, because having devoted my whole adult life to education, I am very interested in rigor and standards and good expectations for students to live up to. But I also don't want teaching—and we have heard some testimony about this, and I think Senator Kassebaum was raising this question—I also don't want this to drive the teaching. I mean there is enough worksheets in the classrooms.

I don't want the teaching to be driven by the tests, because for my own part, while I see education key to the economy, I also,

much like John Dewey, think that we have to have men and women who can think on their own two feet, who can ask critical questions. And I don't want that to be lost sight of. I guess I just want to express that as a concern here.

My question is as follows, and I could be way off base, but this is the one question that haunts me more than any other as I listened to you all today. All of your testimony I appreciate. We heard, I think it was Mr. Anderson, talk about a Marshall Plan. I don't see any Marshall Plan for children in this country. I don't see it in the current budget. I don't see any commitment to anything close to full funding for Head Start. I don't see any commitment to full funding for the WIC program. I don't see any commitment to the idea that every woman expecting a child is going to have a diet rich in vitamin, mineral, and protein.

And I think the medical evidence is unambiguous,

Mr. Chairman, that that has to happen for that boy or girl to have a chance.

I don't see the commitment to early childhood education. I don't see the commitment at the elementary school level, which as a college teacher I would insist is far more important than college teaching in terms of a really less of a ratio of teacher to student.

This is my concern, that in the absence of that commitment, I am worried about some kind of standard testing or moving in that direction becoming kind of punitive. I mean, if we don't really change the life circumstances, you know what I am saying, we don't change the life circumstances of the kids, we don't really make commitments in certain areas where we know darned well it would make a difference, but then we've got these tests along the way or whatever we're talking about, and it becomes kind of tracking. It becomes punitive. It adds, it exasperates the inequalities. That is my concern. I mean I think everybody on this panel wants to see all the rest of that happen. I am not in any way, shape, or form suggesting otherwise. But that is my concern: Where does this fit in, and am I wrong in worrying about this? That is my biggest concern.

Ms. RESNICK. I think you are quite right to worry about it. It is why, in the program that my colleagues and I are working on, which we call the New Standards Project, and you will hear more about from Mr. Tucker later, we insist on the embedding of an examination system in a systemic school reform program. Simply setting examinations will not do anything more than continue to record that we are not doing well.

Senator WELLSTONE. Well, could I quickly interrupt you, Dr. Resnick? I am not asking you to make the decision for me, but I think that would be my position. Do you think that's a credible position: I would support this, but I will not support this unless it is in the context of an overall reform and commitment? What would you do in my place?

Ms. RESNICK. I would take that position and do all the work that is necessary to build the support for it and the funding for it. It will need funds, but it will also need an enormous amount of hard work. It cannot be just the pouring of money into the process.

Senator WELLSTONE. Sure.

Ms. RESNICK. But there will be a need for either new funds or dramatic reallocation of the way funds are now spent in our education and human services system.

Mr. KOLBERG. Senator, I would respond by saying that you are absolutely right, there is no panacea to basic restructuring of American education all the way across the board, and to just do this one little bit and pretend it's a panacea or CHOICE is a panacea just won't get us there.

I think you heard in my testimony I kept talking about national will. I agree we've got to develop a national will. It is not there yet. These kinds of hearings build national will.

But what disturbs me is when you take a poll, 75 percent of the American people say, "Yes, we have a basic education problem." If you go ask parents, "Is the school your child goes to good," the same 75 percent say, "Yes, my child is in a good school."

Parents are confused. We are confused as a society. We talk about the problems, but they don't really reach us. Are American parents really concerned enough about their school? Because if we have a serious national problem, it exists, by and large, in their school as well as everybody else's school. That is why we need these kinds of standards we are talking about, because I don't think parents know that in many schools today their kids aren't getting a world-class education and they don't test that way.

Mr. ANDERSON. Senator, I think you have just simply thrown the cover off of what many of us are really not looking at in a productive way in relationship to educating children down at the local level in the classroom.

Recent statistics tell us that 40 percent of the children of this country are already at risk academically when they enter kindergarten—40 percent. And we are trying to do the best we can, with at-risk programs and a host of other factors.

But let me just give you one example. Tonight I will have a massive meeting in my district in New York in order to listen to the people give me their input on how $13 million, 36 percent of the State aid, based on the governor's $891 million reduction in aid to public schools, can be implemented in my school districts. And what is going to happen, we are going to be back to the very basic core programs and those gifted prime programs, those student support teachers, those off-site schools, and on and on, that we have been implementing to try to lift these kids up by the bootstraps will be eliminated. And we know what will happen to those children.

Yet, at the same time, we try to move forward in terms of setting standards in terms of testing. But you are absolutely correct, there is a greater issue in this country in relationship to education, and if we do not address that issue, we are doomed to generations of kids that are not going to make a meaningful contribution to our society.

Senator WELLSTONE. Mr. Ambach.

Mr. AMBACH. Senator, you asked what we would advise if we were in your shoes. And I state this after hearing your very persuasive statement of commitment to what must be done in this country by way of improving the condition of children and, therefore, the condition for education. If I were in your shoes, I would

insist that any development of a testing system is a testing system that works for each and every individual youngster, not a testing system which screens out but a testing system which in fact measures progress for each and every youngster, to help them progress toward the standards and the objectives that we set.

That is a very bold objective, but I think that the way that you can do that particularly from the perspective of your service here in the Congress is by way of working toward the kind of research and development necessary for an individual examination system, working on the improvements in the NAEP system, so that in fact we are constantly improving the very testing that we are doing.

Right now we are in a very rapidly changing circumstance in this country by way of tests themselves. And that needs a tremendous boost by way of resources so that we are certain we are not doing multiple choice, we are not doing simple identifications, what we are doing in the testing is what is at the core of the very learning itself, and that is, generating new levels of inquiry and questioning by our students so that they are themselves learning, that they are in fact setting their own educational agenda. That is what I would insist on from your perspective.

Senator WELLSTONE. Thank you.

Mr. Chairman, I think I have taken up my time. I would like to apologize.

I would like to thank the panelists from my own point of view, and I would like to apologize to others who are going to testify. I am rather frustrated because I see some of the people that are going to be testifying, and I want to hear them. At least it seems like a kind of appropriate excuse. I made a commitment to go speak to students right now. So I am going to have to do that.

Thank you.

Senator PELL. Thank you very much.

We now come to our second panel. I would add that we have three panels with us today, so we will have to move along as quickly as we can.

Dr. Richard DiPatri, vice president of Educate America, Inc.; Dr. Monty Neill, associate director of FairTest, National Center for Fair and Open Testing; Dr. C.L. Hutchins, executive director, Mid-Continent Regional Educational Laboratory; and Albert Shanker, president, American Federation of Teachers.

We welcome all of you, and as I say, I apologize for the briefness of time that we have, but I am very glad that you are here, and we will start out hearing Dr. DiPatri.

STATEMENTS OF RICHARD A. DiPATRI, VICE PRESIDENT, EDUCATE AMERICA, INC., MORRISTOWN, NJ; MONTY NEILL, ASSOCIATE DIRECTOR, FAIRTEST (NATIONAL CENTER FOR FAIR AND OPEN TESTING), CAMBRIDGE, MA; ALBERT SHANKER, PRESIDENT, AMERICAN FEDERATION OF TEACHERS, WASHINGTON, DC; AND C.L. HUTCHINS, EXECUTIVE DIRECTOR, MID-CONTINENT REGIONAL EDUCATIONAL LABORATORY, AURORA, CO

Mr. DiPATRI. Mr. Chairman, members of the committee, I am truly honored to be here this morning. On behalf of former New Jersey Governor Thomas Kean, who is chairman of Educate Amer-

ica, I thank you for the opportunity to present Educate America's proposal for a national achievement examination.

Educate America's proposal calls for a national achievement exam for all students, which would measure outcomes in six areas: Reading, writing, mathematics, science, American and world history, and geography. The examination would be a requirement for all students in public and nonpublic schools. It would be administered in November of the senior year and would not be a requirement for graduation.

The individual student results would be reported on a numerical scale of zero to 200 in each area assessed. Group results for States and individual schools would be made public, allowing educators and policymakers to focus attention on clear, unambiguous, easy-to-understand results. By so doing, Educate America's proposal would provide a strong measure of accountability for States and schools and specifically for individual students.

Three of the six national goals focus on student achievement and set very ambitious targets for our Nation's schools and States. Unfortunately, and unquestionably, there is no accurate way today for individual schools and States to measure progress toward the national goals. An examination for all seniors would directly link the goals to every high school student in the country.

If it is appropriate to establish national goals, it is even more appropriate to ensure that the goals are met and imperative that these goals affect every student in this country.

In addition to assessing national goals, a national achievement exam would encourage students to work harder in school because the results would provide a common, reliable indicator of their performance, which would have an effect on future job opportunities. By reporting results in an easily understood format which can be shared with employers and postsecondary institutions, the exam would add meaning to the work of students.

More importantly, it would create an effort-oriented system which would send a message to all students that hard work pays dividends and that tough courses are the path to success.

Students would not be alone in being held accountable. By making results public for schools and States, the $230 billion education enterprise finally would be accountable for results, because the results would be comparable across States and schools, and decision-makers at all levels would have sound information to determine where changes are necessary. For the first time in our country's history, a reliable, commonly acceptable indicator of accountability would be available for every high school in the country, and depending upon the results of this objective indicator we can either celebrate success or focus resources where they are needed the most.

When bold proposals such as ducate America's are presented, more often than not they are misrepresented and misunderstood. I would like to take the opportunity to describe to you the highlights of our proposal.

The proposed exam would utilize quality assessment practices. Most of the tests used today result in poor teaching practices and irrelevant curriculum by focusing on minimums or by relying on simple multiple-choice tests. Educate America's proposal would uti-

lize state-of-the-art assessment, including performance measures as often as possible, including essays or a series of essays, techniques to measure ability to solve multistep problems and open-ended questions to measure the ability to think critically and write effectively.

The technology is available today to construct and administer tests which can effectively measure higher-order skills which all educators agree are sound outcomes of schooling. Although more expensive than common testing methods, it is possible and practical to develop high-quality assessment instruments which can be administered on a large scale to produce accurate and reliable information.

In addition to linking curriculum with assessment and motivating particularly noncollege-bound students, Educate America's proposal would raise academic rigor and expectations for all students by administering an exam in grade 12 at the culmination of schooling. While the exam would not be a graduation requirement, it would assess what Andrew Porter of the Wisconsin Center for Education Research calls "hard content" for all students, which contrasts significantly with most State graduation tests which assess "easy content" for all students. This dramatic shift to greater academic rigor and higher expectations for all students is essential if we expect our students to function effectively in the 21st Century.

In conclusion, we believe that this proposal would have the power of fixing the ends and clearly displaying the results for all to see. The results would be fair, clear, direct, and easily understood by all, thereby providing a much-needed focus on results.

On behalf of Educate America and its supporters, I encourage you to give serious consideration to a required national examination for all students. Anything less than an exam which holds everyone accountable—students, schools, and States—will leave you asking the same question in the year 2000 that you are asking today: How can we end this educational crisis in our midst?

Thank you for your attention.

[The prepared statement of Mr. DiPatri follows:]

PREPARED STATEMENT OF MR. DIPATRI

Mr. Chairman, members of the committee, my name is Rich DiPatri, vice president of Educate America, and I am honored to be here this morning. On behalf of former New Jersey Governor Thomas Kean, Chairman of Educate America, and Saul Cooperman, president, I thank you for the opportunity to present our proposal for a National Achievement Examination.

Educate America's proposal calls for a national achievement examination for all students which would measure outcomes in six specific areas: Reading, writing, mathematics, science, American and world history, and geography. The examination would be a requirement for all students in public and nonpublic schools; would be administered in November of the senior year; and would not be a requirement for graduation.

The individual student results would be reported on a numerical scale of 0-200 in each area assessed. Group results for states and individual schools would be made public, allowing educators and policymakers to focus attention on clear, inambiguous, easy to understand results. By so doing, educate America's proposal would provide a strong measure of accountability for states and schools and specifically for individual students.

National Goals Would Affect Every Student

Three of the six national goals focus on student achievement and set very ambitious targets for our Nation's schools and states. Unfortunately, and unquestionably, there is no accurate way today for individual schools and states to measure progress toward the national goals. An examination for all seniors would directly link the goals to every high school student in the country. If it is appropriate to establish national goals, it is even more appropriate to ensure that the goals are met and imperative that those goals affect every student in this country.

Would Create an Effort-Oriented System

In addition to assessing our national goals, a national achievement examination would encourage students to work harder in school because the results would provide a common, reliable indicator of their school performance which would have an effect on future job opportunities. By reporting results in an easily understood format, which can be shared with employers and postsecondary institutions, the exam would add meaning to the work of students. More importantly, it would create an effort oriented system which would send a message to all students that hard work pays dividends and that tough courses are the path to success.

Accountability for Schools and States

Students would not be alone in being held accountable. By making results public for schools and states, the 230 billion dollar education enterprise finally would be accountable for results. Because the results would be comparable across schools and states, decisionmakers at all levels would have sound information to determine where changes are necessary. For the first time in our country's history, a reliable, commonly acceptable indicator of accountability would be available for every high school in the country. Depending upon the results of this objective indicator, we can either celebrate success or focus resources where they are needed the most.

When bold proposals such as educate America's are presented, more often than not, they are misrepresented and misunderstood. I would like to take the opportunity to describe to you the highlights of our proposal

Quality Assessment Practices

The proposed examination would utilize quality assessment practices. Most of the tests used today result in poor teaching practices and irrelevant curriculum by focusing on minimums or by relying on simple multiple choice tests. Educate America's proposal would utilize state of the art assessment practices including performance measures as often as possible, including a series of essays to determine writing performance, techniques to measure ability to solve multi-step problems, and open ended questions to measure the ability to think critically and write effectively. The technology is available today to construct and administer tests which can effectively measure knowledge and higher order skills which all educators agree are sound outcomes of schooling. Although more expensive than common testing methods, it is possible and practical to develop high quality assessment instruments which can be administered on a large scale to produce accurate and reliable information about students.

Link Curriculum With Assessment

The proposed examination would provide an essential link, which does not currently exist, between what students are expected to learn, and the tests used to assess students. Educate America's proposal calls for committees of outstanding educators, business and labor representatives, military officials, and postsecondary educators to identify the common core of knowledge and skills which all students should possess upon completion of secondary school. These committees would define the precise outcomes in each of the six subject areas which would reflect what every citizen, regardless of ethnic background, family income or school of attendance should know in order to participate actively in a democratic society. We live in a highly mobile society which demands that all students acquire essential knowledge and skills to function effectively in that society. The result will be improved curricula with extensive local flexibility, because local educators will choose the best methods, textbooks and instructional materials to achieve the desired outcomes.

Would Motivate Noncollege Bound Students

At the present time, there is no reliable standardized assessment utilized by employers for noncollege bound students seeking employment. The result is the reliance on the diploma by employers for job entry decisions. Educate America's proposal would motivate noncollege bound students because it would supplement students' transcripts and add meaning to the high school diploma. Individual student results would provide to potential employers and other recipients chosen by the student, an easy to understand description of performance in key areas of study. In effect, it would produce a much needed relationship between school performance and employment.

Academic Rigor and "Hard Content for all Students"

Educate America's proposal would raise academic rigor and expectations for all students by administering an exam in grade twelve, at the culmination of schooling. While the exam would not be a graduation requirement, it would assess what Andrew Porter of the Wisconsin Center for Education Research calls, "hard content for all students", which contrasts significantly with most state graduation tests, which assess "easy content for all students". This dramatic shift to greater academic rigor and higher expectations for all students is essential if we expect our students to function effectively in the 21st Century. It is not realistic to suggest that simultaneously we can raise the level of expectations for our students, pegged at the highest international levels, while at the same time testing them when they are in ninth or tenth grade.

Results Easily Understood by Every American

The results for this test would be meaningful and would be understood by every American. This will be an extremely positive turn of events, since we could stop arguing about the extent of the problem and focus on meaningful change. We would then know which programs are working and which are not, and we would no longer rely on the often reported, but least meaningful measure of our school's progress, the sat scores. Despite the fact that the sat is a "prospective test", which is used to assess ability to perform college work, and which is specifically designed to be unrelated directly to curricula in our high schools, the general public annually determines the success or failure of education based on these results. Doesn't it make more sense to use a "retrospective exam" focusing on achievement an performance, which will clearly indicate what students know and how they perform in relation to their school experience?

In conclusion, we believe that this proposal would have the power of fixing the ends, and clearly displaying the results for all to see. The results will be fair, clear, direct and easily understood by all, thereby providing a much needed focus on results.

A Requirement for All Students

On behalf of Educate America and its supporters, I encourage you to give serious consideration to a required national examination for all students. Anything less than an exam which holds everyone accountable—students, schools and states—will leave you asking the same question in the year 2000 that you are asking today. How can we end this educational crisis in our midst? Thank you for your attention.

Senator Bingaman [presiding]. Thank you very much.

Dr. Neill.

Mr. Neill. Senator Bingaman, Senator Kassebaum, thank you very much. FairTest is very pleased to be here today at this important hearing.

FairTest is the Nation's only advocacy and public information organization whose sole purpose is to try to make assessment in the United States air, open, and relevant. It is our conclusion, based on looking at the published proposals, that the Senate at this time should support education reform by not implementing the national examination. The Senate should, however, support efforts to introduce new assessment methods across the country as part of implementing school reform.

42-921 0 - 91 - 3

National testing proposals assume that more measurement will produce positive change. Recent history shows this is not true. During the 1980's the U.S. school children became the most overtested school children in the world, but the desired improvements did not occur. Adding more testing is clearly not the way to improve education any more than taking the temperature of a patient more often will bring down the fever.

Successful educational reform must begin by defining the kind of education we want our children to have. On that basis, we can determine how to make the changes in curriculum, instruction, school governance, and assessment required to reach our educational goals, and then we can decide whether we should have a national exam.

Current proposals for a national test, however, seek to test before the necessary decisions about the goals and methods of school reform have been made. This could lead to the imposition, without adequate public discussion, of a national curriculum and a national school board that would have the effects of eroding democratic control of education and local accountability and will end up short-circuiting the developing processes of school reform.

The national examination systems of many other countries are based in school systems that rigidly sort their students and perpetuate social class divisions. No one test should become a national gatekeeper that continues our Nation's unfortunate history of unfairly sorting children by race and class.

Some current proposals, such as that of Educate America, call for the creation of a low-cost test that would be administered to all students in the near future. Because of cost and time factors, it is our conclusion that such a test inevitably will be almost entirely multiple-choice. Multiple-choice tests have already been criticized. I would just like to repeat they do not provide much useful information for instruction or for policymakers, they cannot assess higher-order thinking capabilities, and, therefore, they won't give us any information about that. They will reinforce incorrect ideas about how people learn and will, therefore, reinforce poor educational practice. Teachers will teach to it if it is a national test, and instruction will be reduced to drilling for multiple-choice exams.

In short, no proposal that will rely on or lead to national multiple-choice testing should be given any serious consideration by the Senate.

Implementing a national performance-based examination on the other hand does have advantages over multiple-choice testing of using methods that can assess higher-order abilities and that support good educational practice. However, it is FairTest's conclusion that we can move toward the use of such assessments without constructing a national examination system. We then gain the advantages of good assessment and avoid the disadvantages of imposing a national testing program.

Assessment should be one part of reform, not the controlling force. If we focus on assessment as the solution to our educational problems, we may well fail to address such critical issues as equity, rigid and bureaucratic school governance and structure, low-quality textbooks, inadequate schools of education, and a lack of information about school inputs, processes, programs, and other outcomes.

We would also be imposing new exams and tests on teachers who won't be able to use them without adequate education themselves.

In short, implementing a national performance-based exam at this time puts the cart of testing before the horse of systemic educational reform.

We also do not know yet whether it's really feasible to construct such a national system of exams. The whole process, particularly the calibration, could prove too complex, expensive, and unwieldy to work.

Britain was mentioned by Mr. Kolberg. They have dropped the moderation part of their national examination system. Moderation is the means by which you constantly recalibrate what students are able to do, what good work is, how teachers can grade these papers. They dropped it, my understanding is, because it was too expensive. Moderation would be essential to any good national examination system. It is our fear that when the complexities become clear, that the complex performance-based exam will be dropped and we will end back toward multiple-choice testing.

Our recommendation to the Senate is that instead of implementing a national examination at this time, that you work to assist the States and the districts, acting in consortia, to develop and implement performance-based methods of assessment, that you assist State and district teacher education and staff development programs, that you assist the subject-area groups to develop and disseminate model curricular standards and assessments, that you ensure that the national assessment of educational progress remains simply and only a national monitoring system. And only after these education reform processes have been implemented and evaluated should the Federal Government consider whether it is desirable or feasible to link the newly developed State and local performance-based assessments to each other and to national standards or curricular frameworks.

Thank you very much.

Senator PELL [presiding]. Thank you very much, Dr. Neill.

[The prepared statement of Mr. Neill follows:]

FairTest
National Center for Fair & Open Testing

Why the United States Does Not Need a National Test:

Testimony to the Senate Committee on Education

by

Monty Neill, Ed.D., Associate Director,

National Center for Fair & Open Testing (FairTest), Cambridge, Mass.

March 7, 1991

Ladies and Gentlemen,

Thank you very much for inviting FairTest to appear at this important hearing.

Based on an examination of existing proposals, FairTest concludes that most current efforts to establish a national test to measure progress toward the nation's educational goals will hurt, not help, our nation's efforts to improve school quality. We therefore urge the Senate to support education reform by ***not*** implementing a national exam at this time. The Senate should, however, support efforts to introduce new assessment methods as part of implementing school reform.

National testing proposals largely are based on the false premise that measurement by itself will produce positive change. Recent history shows this is not true: During the 1980s, U.S. school children became probably the most over-tested students in the world -- but the desired educational improvements did not occur. FairTest research indicates that our schools now give more than 200 million standardized exams each year and the typical student must take several dozen before graduating.[1] Adding more testing will no more improve education than taking the temperature of a patient more often will reduce his or her fever.

In contrast, successful educational reform must include restructuring curriculum, instruction, textbooks and other materials, school governance, and teacher education, as well as assessment. What we need to create are schools as communities of and for learning.

To move toward that goal, teachers, administrators, other school personnel, parents, students, community members, and government must all be involved in an open and democratic process of defining our educational goals -- at the local, state and national levels -- so that we can agree, for example, on what it means for all students to be competent in different areas. On that basis, we can then determine how to make the changes required to reach the goals, including a decision on whether to institute a national test. Most current proposals for a national test, however, seek to test *before* necessary decisions about the goals of school reform have been made. This likely will lead to the backdoor imposition of a national curriculum, without public discussion.

Indeed, having a single national test raises the issue of the control of education. If the test becomes important, as all testing proponents want, those who control the test could control curriculum and instruction, particularly if decisions about curriculum and instruction

342 Broadway, Cambridge, Mass. 02139 (617) 864-4810 FAX (617) 497-2224

have not been arrived at before the test is constructed, and maybe even if those decisions have been reached.

A national exam should not be allowed to undermine such needed and emerging reforms as school-based management and shared decision-making. By centralizing decision-making, centralized national testing most likely will make education less, not more, accountable to parents, students, teachers and the community. If the test is centrally controlled, to whom could parents, teachers and communities appeal if they disapprove of the curricular decisions and instructional methods imposed through the test?

Any national test or examination system could produce additional major problems. For example, the examination systems of many other nations are based in school systems that rigidly sort their students and perpetuate social class divisions, precluding the social mobility our nation believes in. No one test should become a national gatekeeper that perpetuates our nation's unfortunate history of unfairly sorting students by race and class.

A national test could end up being used to determine high school graduation, employment and entrance into higher education. Due to unavoidable measurement error and bias, many students who fail a test will, in reality, be as capable as many who pass. Research indicates that those who fail but should have passed will be disproportionately from low-income and minority-group backgrounds. FairTest agrees with the National Commission on Testing and Public Policy that, because of the bias and error, no one test should ever be the sole or primary basis for making an important educational decision.[2]

Dangers of Multiple-Choice Testing

FairTest recognizes that there are two different types of proposals for national testing. One type will rely essentially on multiple-choice testing; the other calls for performance-based assessment. These two approaches are quite different. They are the difference between testing *what students should know* and *what students know how to do*.[3]

The first approach quickly leads to multiple-choice testing of arbitrary facts and isolated skills, unconnected to the way knowledge is used in the world. Multiple-choice and short answer tests cannot adequately assess problem-solving or the ability to create and use knowledge.[4] Higher order thinking requires the student to define the problem, to consider and attempt various solutions to problems which are ill-structured and may have more than one correct solution, and to produce knowledge, not merely recognize answers.

Because multiple-choice/short answer testing cannot directly assess higher order capabilities, a test comprised of such items will not inform us as to the problem-solving and knowledge-creating capabilities of our students. We know from research, however, that student abilities in these areas are very limited. This has been caused largely because schools of schools' failure to teach them in any subject area to more than a few students. Even the best high school students typically do not know how to problem-solve using the approaches and methods a professional uses.[5] Yet research also shows that problem-solving, knowledge-creating approaches can be used even with very young children.[6]

If a test is important -- as a national test is sure to be -- then teachers will teach to it.[7] Because multiple-choice tests cannot directly measure higher-order skills, teaching to the test reduces or eliminates instructional time spent on the higher skills. Instruction is reduced to drilling for multiple-choice exams and the curriculum is reduced to the test. Multiple-choice

testing precludes a curriculum based on thinking, investigating, problem-solving and using creativity, because the test cannot measure those things.

Additionally, these tests perpetuate the false idea that first students learn basic skills, then they learn higher skills. Cognitive psychological research has demonstrated that learning involves active thinking and to enhance learning the student must be actively engaged.[8] Test-driven schools produce higher test scores, but not students who are able to think.

A predominantly multiple-choice test may include a writing sample. A typical short writing sample requires a student to write several hundred words on a topic he or she may or may not know anything about and may or may not care about, in a short period of time, with no chance for research, discussion (that is called cheating), or serious revision, for no purpose except the test. If the purpose of writing is to communicate, then a typical test writing sample cannot legitimately be called writing at all. As with multiple-choice testing, it sends the wrong message about the goals of education.

These tests are not very useful to teachers or policymakers. The reason, in both cases, is that the test results do not help the teacher or policymaker decide what to do next. If Johnny cannot multiply, the test cannot explain why. If Maria's whole class cannot multiply, the test does not provide information on what should be done.

What standardized multiple-choice tests do best is help sort students: the good, the bad and the ugly. It is what they were invented to do. But if we are serious about reforming education so that all students can learn the things we deem important, then we must stop relying on tests that have as their only real use the sorting of students.

In sum, implementing a national multiple-choice exam will mislead the public about the nature of the problem and the requirements of real change, block positive school reform (including the use of new methods of assessment), hinder students' ability to develop the kinds of intellectual competencies they need to develop, and ultimately undermine public education.

No proposal that suggests using more than a small proportion of multiple-choice items in a national examination should be given any serious consideration by the Senate. At most, multiple-choice could be used as part of a sampling program to gather limited information about student acquisition of a narrow range of knowledge. There is no reason to test every student for this purpose and such a purpose should never be allowed to dominate education, as it too-often now does.

Because of current technical limitations, any proposal to assess our nation's students inexpensively and in the near future will, of necessity, be a multiple choice test. An example is the proposal by Educate America to test all high school seniors in six subjects for $30 - $50 each. Such proposals must be rejected.

Performance-Based Assessment

By contrast, students should be assessed on what they know how to do. To know how to do something includes knowing factual content. This method of assessment corresponds to how people learn. They learn by integrating new information or experiences into the intellectual frameworks they already possess, which in turn enables them to refine and improve the frameworks.

Assessing what students know how to do is based on students' doing real work. There are many ways for students to demonstrate intellectual competence in and across the subject areas. Performance-based assessments can be based on regular student classroom work - projects, research, writings, products, self-reflection, teacher evaluation, exhibitions, and performances - that can be organized and summarized in portfolios. In turn, the portfolios can be examined by outside people -- teachers, other parents, trained examiners -- to determine the quality of the portfolios and the kinds of work students are doing. Vermont for example, is working on this method.

Performance-based assessments can also be examinations administered from outside the classroom. These can include open-ended, complex problems requiring the student to figure out what to do, solve the problems, and explain what he or she did. Or they can be exhibitions, performances and products, such as now done in science fairs, Scout Merit Badges, Advanced Placement Art, and many performing and applied arts. These often can be exams that are worth teaching to, unlike multiple-choice tests. Arizona, California, Connecticut and Maryland are among the states implementing these types of exams.

Taken together, in-class and externally-developed performance-based exams can encourage real work, model high standards, spur improvements in teaching and curriculum, produce instructionally useful information for teachers and students, and provide information based on real activities about student progress. Assessment can play an important part in developing communities of and for learning.

Cautions on a National Performance-Based Examination System

However, support for performance-based assessments does not mean such assessment should immediately be transformed into a national examination system, such as that proposed by the Learning Research Development Center and the National Center for Education and the Economy (LRDC/NCEE).[9] There are many reasons why this is the case. Among them are:

-- We have not yet completed the process of discussing and debating what we want our educational systems to be. Many complex issues of educational reform, involving curricular goals and standards, instructional methods, assessment methods, school structure and governance, and collection of information, largely must be resolved *before* the question of whether a national examination system is desirable can be answered. To do otherwise is to put the cart before the horse.

-- Imposing a national examination will not address the issues of rigid and bureaucratic school governance and structure, low-quality textbooks, and inadequate schools of education. Improving assessment needs to be considered as one part of integrated systemic change.

-- The proposal calls for national boards to set standards. It could create a national school board that, by setting curriculum standards, will lead to a centralized, national education system. Because the consequences of such actions cannot now be known, but may include undermining democratic control of education, we should not rush into that process.

-- Staff development is central to school reform, but is not adequately addressed in the LRDC/NCEE proposal. If teachers are to teach to performance-based assessments, to teach the "thinking curriculum," they have to know how to do so. This involves developing the ability of our nation's 2-1/2 million teachers to teach and assess in new ways. To be

effective, school reform must include the active participation of those who will implement the changes. We cannot impose new assessments on teachers, change nothing else, and say "Do it."

-- In general, the proposal does not adequately address equity issues that must be solved for the system to be fair. Changing assessment will not by itself reduce inequities. All students must be assured a fair opportunity to learn how to work within a thinking curriculum that uses performance-based assessments. Additionally, the goal of "initial mastery" could encourage sorting and tracking students according to who can best or most quickly reach the goal. This danger needs to be seriously addressed to try to ensure structures and processes, including in the realm of assessment, that are inclusive and reduce tracking and other kinds of sorting.

-- We simply do not know whether it is feasible to construct a national examination system. The whole process, particularly the calibration, could prove to be too complex, expensive and unwieldy to work. For example, England recently dropped a moderation process from its national exam process because it was too expensive. Moderation is the process by which teachers help shape standards and learn to grade papers, products and performances uniformly so as to produce consistent and reliable results. Moderation is valuable and necessary and must be included in any performance-based system, but doing it on a national level on top of state and local levels may be too much as well as unnecessary for educational improvement.

-- When the complexities and expense of the proposal become clear, the portfolios and projects could end up being reduced to very limited exams. There even could be a return to multiple-choice and short-answer exams. Such a retreat would have destructive curricular effects and undermine all aspects of educational improvement.

-- The proposal is not conceived of as one part of an overall educational information system. Having assessment outcome information on education is not useful unless we also have adequate information on inputs (money, teaching staff, building quality, etc.), processes and programs (curriculum, instructional methods, textbooks and materials, class size, role of tracking, governance and school organizational structure, etc.), and additional outcome data (employment and further education of graduates, dropout rates, etc.). This information should be obtained without harming education -- unlike what has happened with multiple-choice tests. Schools and programs should be evaluated on a comprehensive range of indicators of their quality as communities that support learning for all students.

Recommendations

There is no one, simple method of putting a national education reform process into motion in the right direction. It is a process that can and is happening at all levels: the classroom, the school, the district, the state, consortia that include all of these, and at the national level. It is not and will not be a smooth and easy process. But as good practice becomes available to replicate, as improved curriculum and assessments become more widely known, as our nation's desire to improve education for all continues to grow, then we can expect to see real progress.

The federal government can proceed in one of two ways. It can impose a national test that runs the risk of short-circuiting the process of school reform. Or it can find ways to support school reform activities without imposing a national test.

FairTest concludes that the Senate should not propose a national exam either immediately or to be in place within any fixed timetable, such as five or ten years. Rather, FairTest urges the federal government to take the following steps to improve education and assessment:

-- Assist states and districts, acting in consortia, to develop and implement performance-based methods of assessment.

-- Assist state and district teacher education and staff development programs.

-- Assist the subject area groups, such as those in math, English, social studies and science, to develop and disseminate model curricula, standards and assessments.

-- Re-examine the instances in which the federal government requires standardized multiple-choice testing, particularly for the Chapter I program. The testing requirements virtually force programs into being test-coaching programs, though that, as explained above, is a poor educational method.

-- Leave the National Assessment of Educational Progress as a national indicator.[10] To turn it into some kind of a national test will end up destroying its current usefulness and may produce all the drawbacks discussed above. In particular, NAEP should not be used below the level of state-level comparisons. FairTest doubts that state-level comparisons will be of real use to educators and urges that state comparisons not be approved beyond trial measures unless experience and research demonstrate how the comparisons will be used to improve education. NAEP should, however, include far more performance-based assessments and provide technical assistance to districts, states and consortia who are implementing performance-based assessment.

-- Consider how assessment information can best be included as one element of school reform activities and one part of an indicator system, and not view assessment in isolation.

Only after these educational reform processes have been implemented and evaluated over a period of time should the federal government consider whether it is desirable or feasible to link the newly developed local and state performance-based assessments to each other and to national standards or curricular frameworks.

Let us be clear. FairTest is are not arguing against accountability or for slowing down school reform. Nor is the issue one of the need for "standards." Rather, the central issue is how we define education. We are saying that we need school reform, not more testing. We need genuine accountability, not test scores from multiple-choice or short-answer exams, and we don't need to jump aboard an examination train heading into trackless terrain.

Our nation must not be misled into thinking more testing will solve our educational problems. Instead, we must construct plans for reform that include assessments which can be used to help student learning, guide educational improvement, provide information for accountability, and assist the goal of equity, but not block progress or harm students. Our nation will be far better served to take the time to do the job well, than to act hastily and poorly with destructive results.

Endnotes

1. Medina, Noe and D. Monty Neill, ***Fallout from the Testing Explosion: How 100 Million Standardized Exams Undermine Equity and Excellence in America's Public Schools*** (Cambridge, MA: FairTest, 3rd Ed., 1990).

2. National Commission on Testing and Public Policy, ***From Gatekeeper to Gateway: Transforming Testing in America*** (Chestnut Hill, MA: Author, 1990);

3. Susan Harman, drawing particularly on the work of people associated with the Coalition for Essential Education, uses this formulation very clearly in "National Tests, National Standards, National Curriculum," ***Language Arts*** (January 1991: 49-50).

4. For a general critique of multiple choice testing, see Neill, D. Monty and Noe J. Medina, "Standardized Testing: Harmful to Educational Health," ***Phi Delta Kappan*** (May 1989: pp. 688-697), and Medina and Neill, *op. cit.* For analysis of why multiple-choice testing cannot assess higher order thinking, c.f. Fredericksen, Norman, "The Real Test Bias: Influences of Testing on Teaching and Learning," ***American Psychologist*** (March 1984: pp. 193-202); National Commission on Testing and Public Policy, *op. cit.;* Resnick, Lauren B. and Daniel Resnick, "Assessing the Thinking Curriculum: New Tools for Educational Reform," in B.R. Gifford and M.C. O'Connor, eds., ***Future Assessments: Changing Views of Aptitude, Achievement, and Instruction*** (Boston: Kluwer Academic, 1989).

5. "Beyond the Bubble," a mini-conference at the April 1990 national conference of the American Educational Research Association, Boston, MA.

6. Kamii, Constance and Mieko Kamii, "Why Achievement Testing Should Stop," and Engel, Brenda S., "An Approach to Assessment in Early Literacy," in Kamii, Constance, ed., ***Achievement Testing in the Early Grades: The Games Grown-Ups Play*** (Washington, DC: National Association for the Education of Young Children, 1990).

7. Madaus, George. "The Influence of Testing on the Curriculum." *87th Yearbook of the National Society for the Study of Education, Part I: Critical Issues in the Curriculum* (1988: pp. 83-121). Also, Resnick and Resnick, *op. cit.*

8. Resnick and Resnick, *op.cit.*

9. Learning Research and Development Center and National Center on Education and the Economy. *Setting a New Standard: Toward an Examination System for the United States, A Proposal* (Pittsburgh and Rochester: Author, October 1990) is the most comprehensive of the reports and documents proposing a national performance-based examination system.

10. Campaign for Genuine Accountability in Education, "Open Letter to Congress, Bush Administration, the Governors on NAGB and NAEP Expansion," (Cambridge, MA: FairTest. June 15, 1990).

Senator PELL. Mr. Shanker, an old friend of the committee. Welcome back.

Mr. SHANKER. Thank you very much, Mr. Chairman and members of the committee. I have a written statement, which I won't read.

I would like to place this discussion in context.

Senator PELL. Your full statement will be in the record.

Mr. SHANKER. Yes. Thank you.

I think we have to face the fact that we are in a State of crisis with respect to public education in America. There is a very great, rapid loss of confidence, and in terms of performances, ability of our students to perform, we are far, far below what other major industrial countries are able to do. If we just look at the NAEP results, we see that, at age 17, those of our youngsters who are still in school after 25 to 35 percent have dropped out, only between 3 to 6 percent are able to write a good letter or essay or are able to do some problems in elementary algebra or two-step arithmetic problems or are able to read anything that has any complexity, things that are worthwhile.

If you compare that with the percentage of youngsters who pass college entrance examinations like the Arbiture in Germany or comparable examinations in France or Great Britain, other countries are producing between 16 and 28 percent of the graduates who are able to function at a level equal to or higher than that of our top 3 percent.

Now, it used to be the story that those are elitist systems and that they only educate their top kids and somehow they educationally throw the others away. That is not true. The bottom track in Germany provides a very good education, one which we would be happy to have most of our kids get and to achieve at that level. The same is true, with the exception of Great Britain, which does have a very serious problem with its bottom kids. All the other industrial countries do not sacrifice the others for the sake of those at the top.

Now, I think we need to ask ourselves, many of the arguments that we hear, there is no question that there are all sorts of difficulties with this issue and there are all sorts of things that can go wrong with it, and if we weren't facing great difficulties, we could say, well, these problems are so great, maybe we should just wait and think about it a little longer. I think, given our current crisis, that to think about it a little bit longer without thinking about what such a system of examinations could do to strengthen our educational system is dead wrong. It's being over-cautious.

As the chairman has said for many years—and there are more people coming to think your way—that with all the problems that are there, and we've got to watch them as we move ahead, the price that we pay for not having a system of examinations, national examinations, is a tremendous price. We have 15,000 separate school districts all defining what it is that kids ought to know and be able to do. That is absolute idiocy and disaster.

And we also have a teacher training program which is unique in the world because if you haven't defined what kids ought to know and be able to do, you have no way of deciding what should teachers learn, because in every other country what teachers are trained

to do is to teach the curriculum. And if you don't have a curriculum, you have no basis, really, of intelligently selecting and training teachers.

So, we've got a system in which we've got these standardized multiple-choice tests, which you don't study for. Now, if there is anything worse than a test that you do study for it's one that you don't study for.

There is nothing wrong with teaching to the test if it's a good test. There is nothing wrong with having your teachers talk about how do you organize an essay and how do you present a persuasive argument and how do you amass evidence and how do you question counter-arguments. There is nothing wrong with studying for a test like that and with teachers teaching to a test. So it's not a question of that. We should have tests that teachers can and will teach to because they're good tests.

And I don't think we should avoid our responsibilities by saying we shouldn't have a system of tests because maybe they will be bad. Well, what if they're good? Even if we start with some that aren't so good, if you've got them, you can improve them. You could have some of the very critics who are here come back each year and tell you how to make them better. That's a lot better than talking about whether you should have them at all.

Now, I want to deal with the issue of whether tests in and of themselves will do anything. In and of themselves they won't. I think it's important that we face the issue that tests and doing well on them will only do something if there are consequences, and I think the fact that so many kids do so well in these other countries is largely because if you don't pass the test you don't go to college. Or if you do get a high results on a test, you get a better job or you get a faster job or you're eligible for certain things.

Youngsters are just like all other human beings: We do some things because we are intrinsically interested and motivated, and we do other things because we want to get somewhere and that's why we sometimes do some unpleasant and painful things. And we have largely become a society where what you do in school doesn't make any difference. You can get into college whether or not you've learned anything, and you can get out of college and get a degree whether or not you've learned anything.

We have got a lot of outstanding employers who won't hire any 18-year-olds; they wait till they're 24 or 25, which means that every 18-year-old gets a lousy job no matter how well he's done in school.

So I am here not only to argue that we need to move forward toward a system of national examinations but we should not shy away from the notion that how well one does on these examinations, assuming that they are good and adequate measures of what we're doing and we're doing things that are sensible, that there ought to be consequences attached. And it's the attachment of consequences that moves both the adults and the youngsters in the system to do better.

[The prepared statement of Mr. Shanker follows:]

Prepared Statement of Mr. Shanker

On behalf of the 750,000 members of the American Federation of Teachers, I appreciate the opportunity to appear before this distinguished committee to address the issue of national standards and testing in American education.

As recently as a few years ago, anyone who even dared to talk about national standards and a national examination system for U.S. schools would have been considered a radical or risk-loving person. Sure, that's the way they do it in most other industrialized countries, he would have heard, and, sure, their students achieve at a much higher level than ours. But the education systems in those countries are under the control of their central governments, and the idea of our Federal Government dictating what children learn in local schools is out of the question.

Now, however, we are beginning to understand something that Senator Pell understood some time ago when he first proposed national exams: We pay a heavy price for our fragmented system—or non-system—of education standards and testing. And now, too, we also are beginning to understand that there are ways of building a national system of standards and examinations in a typically American way that does not involve Federal control over our schools.

Why should we be so eager for national standards and examinations? Exactly what difference do they make in an education system—and, ultimately, in what children learn?

National standards in education mean that there is agreement about what students ought to know and be able to do and, often, about the age or grade at which they should be able to accomplish these goals. Exams based on these standards mean that at any given time, an educator can tell a parent and the public, "Here is what we expect of youngsters in mathematics or biology or composition, and here is how you'll know if our students achieved these goals and how our schools are doing."

National education standards and examinations therefore go hand in hand. They allow students and schools to know what's expected of them, and they give parents and the public a clear means of understanding what our students and schools are supposed to be doing and whether or not they are succeeding. Those are the hallmarks of a comprehensible and fair accountability system, which is something we very much need and have never managed to produce to anyone's satisfaction.

Moreover, most countries that link together national education standards and examinations for students reap the additional benefit of ensuring a better-prepared teaching force because, once you achieve consensus on standards and examinations for students, you have an answer to the question of how to train and assess teachers: Teachers have to be able to teach curricula that reflect the national education standards, and they have to know the various ideas and strategies for teaching such curricula to the diverse youngsters who make up our schools.

In the United States, where we have no such agreement about what students are supposed to learn, we have no linkage between curriculum, the tests used to assess students' knowledge and skills, and what we expect our prospective teachers to know and be able to do. Each of our 15,000 school districts and 50 states has some rights in these areas, and, arguably, so does the Federal Government and the tens of thousands of individual schools in this Nation.

One result are now-you-see-them, now-you-don't education standards that tend to be set to the lowest common denominator. Another result is a bewildering, fragmented, fractious non-system of education where some children might not be exposed to, say, science until secondary school, and where a child moving to another district in the State, let alone another State, could easily feel as if he's just enrolled in a school in a foreign country. And this in a Nation that values equal educational opportunity, whose people move more often than in any other country in the world and whose future depends on being able to meet or exceed world-class education standards! Put another way, while our lack of clear and high standards, and the examinations to support them, may not be wholly responsible for the fact that the performance of our education system is among the lowest in the advanced industrialized world, it is a significant part of the story.

Let's take the example of how other countries test students and how we do it. In most countries with national standards, tests usually consist of writing essays or solving problems based on what the students are supposed to be learning. And when youngsters, with the help of their teachers, prepare for these exams and review questions that were on previous exams, it's a worthwhile educational experience. Writing an essay on the causes of World War I or presenting the arguments for and against imperialism is a good exercise in learning substance and in learning how to organize your thoughts. And the quality of the essay really shows how well the stu-

dent has mastered the material. Countries that also use performance assessments have even more to go on.

But in the United States, we primarily use standardized, multiple-choice tests and use them to test little bits of knowledge that are not directly related to the curriculum. In fact, because curricula vary by state and by school districts within states and even by schools, companies that design standardized, multiple-choice tests are encouraged to and pride themselves on divorcing their tests from curriculum. Moreover, since curriculum-free tests are supposed to be kept in the dark, going over questions from previous tests is considered cheating. It's also a waste of time. Whatever little bits of information the kids do learn have no context, so they'll be forgotten in a hurry. Furthermore, parents looking at their children's test results or someone reading averages scores in the newspaper will have no idea what they represent in terms of what students know or can do. And we have no way of knowing—or at least no one much seems to care—if a school got high scores because it put kids through low-level, multiple-choice-type teaching all year long or because it ignored the pressure to prepare for the tests this way and really educated them. Nor can we tell if low scores mean a lousy school or one that was covering concepts and skills that weren't captured in the test.

Or consider the effect of our lack of education standards on teacher education. When there are no standards for what we want students to know and be able to do, it's certainly hard to agree on what teachers need to know and be able to do. And because their students will end up teaching in many different school districts and many different states, each of which have different curricula, colleges and universities can't train teachers on the basis of the curriculum they are going to teach or assess them on how well they know it. It's little wonder, then, that teacher education is floundering and that most teachers say their training was not helpful in preparing them to teach.

The question is, is it possible to develop a national system of education standards and examinations without the Federal Government actually doing it? Can we reap the benefits of such a system without incurring the risk of Federal control of our schools? I believe the answer to both questions is yes. In fact, there is already evidence that this can be done.

The National Council of Teachers of Mathematics and the Mathematical sciences Education Board have already put together national standards for mathematics that have won widespread support. It also looks like this accomplishment will lead to efforts to develop exams based on these standards. The American Association for the Advancement of Science is far along on its Project 2061, which promises to do the same for science. Teachers and scholars in each field need to follow the lead of these groups and get together to define, with input from the public, what American students ought to know and be able to do. This does not mean devising a single curriculum that prescribes precisely what everybody will learn and how. Nor does it mean developing a single make-or-break test. It means devising curriculum frameworks that reflect the standards we wish students to meet but whose precise content can be set by states, districts, schools and teachers. And it means developing model exams that embody those standards and that students and teachers prepare for naturally as part of the process of teaching and learning.

This process is just beginning, but it looks promising. If it succeeds, we'll have the strength of a national system of standards and examinations without surrendering the freedom to make important choices on the state and local and school levels. And we'll have a revolutionary development in American education carried out in a uniquely American way, a way that is consistent with the values underlying our non-federal school governance system—through the voluntary effort of professional groups and states and, I hope, with Federal support.

The time could not be better to encourage this effort. For the first time since the question has been asked on surveys, a majority of the American public favors the idea of national education standards and examinations. For the first time in the history of our Nation—a Nation whose public education system is central to the strength of its democracy and economy—we have a set of national education goals. These goals were the product of an Education Summit between the President and the Nation's governors, and they have been widely endorsed.

We must now take the next logical and necessary steps. What do these goals actually mean? How can we mobilize our students and schools to achieve them? How will we know if we are making progress toward achieving them? Surely a large part of the answer is to develop a national system of education standards and examinations.

The AFT is prepared to discuss this issue with you further and to provide any supporting materials that could help you in your deliberations. Let me also com-

mend to you a recent article on this set of issues by Marshall S. Smith, David K. Cohen and Jennifer O'Day, "National Curriculum, American Style: What Might It Look Like?," which appears in the Winter 1990 issue of the *American Educator*, the journal of the AFT.

Thank you.

Senator PELL. Thank you very much indeed, Mr. Shanker.

Dr. Hutchins.

Mr. HUTCHINS. Thank you, Mr. Chairman, members of the committee. I am Larry Hutchins, executive director of the Mid-Continent Regional Educational Laboratory, headquartered in Aurora, Colorado. Today I appear before you on behalf of the Council for Educational Research and Development. I have a statement which I wish to enter in the record, and I will just comment on a few highlights.

What I want to do is to express my reservations and the council's reservations about the rush toward the development of a single national achievement test or even, as some people argue, a national curriculum.

Let me start by making my grounds clear. My grounds are the limitations such a test puts on the purposes of education. Today the call for a national test and indeed a national curriculum is justified by a policy that views education as an instrumental function of society; that is, such a test is dependent on the ground that the Nation is falling behind in some international race and we must have higher achievement in order to be No. 1.

While such views may be laudatory, it begs the second educational function, the development of the individual. Throughout our history, education has been the principal tool of freedom, freedom of the individual to think what he or she chooses to and to set goals, whatever they are, that are best for that individual.

I believe that the establishment of a single test or a single national goal for students would ultimately deprive us not only of the diversity on which the Nation was founded, but which sets us apart from all other developed nations.

Don't misunderstand my point. I am not against high standards or setting standards in literacy, numeracy, or scientific knowledge. Public policymakers such as the governors have the right to know how their investment in education is paying off in terms of the fundamentals of human capital. But I believe they should gain that information without unduly constraining the rights of the individual to set his or her own educational goals.

We have the technology to develop assessment procedures that can help us track our progress toward national goals without at the same time setting goals for individuals. That technology is already in place in the form of the sampling procedures used by the National Assessment of Educational Progress, which I might add are being expanded to make the data meaningful at the State level.

The point is simply that we must avoid turning our need for better data about educational progress into goal-setting activities for individuals, which is exactly what will happen if the test becomes an every-student test. Such a test will quickly produce a national curriculum. We must avoid confusing the need for measuring national progress against the desirable standard and using the standard to set individual goals.

The problem is that today's schools are rooted in the 19th Century, and the issue we must face is one of design. Instruction is based on models of lecture, drill, and practice. The model is passive and assumes the child is an empty vessel into which to pump information.

Today's curriculum was developed in the 1890's. It models the reductionist views of Newtonian science, a science being radically changed by physicists and other scientists into a more relativistic, systemic-oriented curriculum.

To develop a national test on the current curriculum would freeze in place a system that is already out of date, and it would have a chilling effect on badly needed innovation.

Consider, for example, the changes we need to bring instructional practice into line with current views of psychology. We know that individuals are not motivated by externally determined goals. Humans are goal-driven, but those goals must be internally set and congruent with our own vision of ourselves.

We won't get more math and science pumped into kids' heads by willing it on them. It will only come about when, first, our schools see their primary function as the empowerment of the individual to set high standards for him or her self; second, when we have designed instruction that is responsive to individual differences and inspires each student to strive for excellence; and third, when we have a new curriculum redesigned around more cognitively complex, systemic, contextually situated issues that are relevant to the 21st Century. These issues include such things as the environment, energy, technology, productivity, space, social justice, communication, issues that are not now central to the existing curriculum. A national test at this time runs the risk of freezing out the very kind of reform that we need in curriculum.

If the design and maintenance of a national test is left to a bureaucratic process, as it inevitably would be, the items on the test would not reflect what we should teach, but what we do teach, thereby freezing the past in place.

Let me put it another way: In the field of physics, knowledge has changed so rapidly in the past decade that any effort we made to drive curricula through a test would be out of date before we could institutionalize the test and report the results.

Let me give you a specific illustration from a State I am working in here there are a large number of Indian children. They are using such concepts as the entrepreneur as warrior, and they have defined one kind of bravery as what a young Native American does when he chooses or she chooses to work outside of the reservation. In this school, the science program is very holistic, focusing on the environment in which these students live, to understand the ecology and nature needed to protect that environment.

I am skeptical of how well these students might do on a national science test, and yet I think they are learning better science that students in many of the suburban schools in my region.

Let me just close with several recommendations. You have already heard that we oppose a single graduation test. We want, instead, to have the Senate spark a redesign of curriculum and instruction, and we have concern that if such a test is established, that there be some way to hold those who instigate it accountable

for its abuses. If the policymakers that establish the test can't guarantee that there won't be abuses, then the test must be accompanied by a process for ensuring that students who take it, or their parents, are given the right for informed consent.

Thank you, Senator.

Senator PELL. Thank you very much, Dr. Hutchins.

[The prepared statement of Mr. Hutchins follows:]

PREPARED STATEMENT OF MR. HUTCHINS

Good morning. I am Larry Hutchins, Executive Director of the Mid-continent Regional Educational Laboratory, headquartered in Aurora, Colorado, just outside of Denver. Today I appear before you on behalf of the Council for Educational Development and Research.

The Council's mission is to support the Congressionally created educational research and development institutions as they find ways—either through their own investigations or their evaluation and use of other research—to enable every American school child to be successful.

Mr. Chairman, improving the performance of our elementary and secondary school students requires many strategies. Some of these strategies involve the Federal Government. We hope that this hearing today will create more public awareness of the need for school improvement and sustained commitment and support for public education. Our comments today are directed toward this goal.

In this testimony, I will express reservations about the rush toward development of a national achievement test or assessment process and, even, as some quarters argue, a national curriculum.

Let me start by making my grounds clear. I don't believe we should object to a national test because it violates the principle of local control—that's like arguing for mediocrity for the sake of representativeness. If, in fact, we are a Nation at risk because of the low level of our educational achievement, local control ought not be used as a rationale for taking dramatic steps to correct the situation. Now, there might be some good constitutional reasons why such a test should not be a Federal test, but one should not argue against a national test if it is, in fact, needed.

Nor do I believe we should object to a national test because such a test might lower expectations and reduce our goals to a minimum standard. As long as parents have some options about where they send their students to school—either by choosing a private school or by deciding where they live—as long as we have academic freedom that lets teachers challenge students in their own way, and as long as we have freedom of thought, one should not be overly concerned that a national test would somehow lead to a leveling of achievement.

But I believe there are other grounds on which to express reservations about the desirability of a national test or assessment procedure.

Those grounds are the limitation such a test puts on the purposes of education. Today, the call for a national test and, indeed, a national curriculum, is justified by a policy that views education as an instrumental function of society. That is, such a test is defended on the grounds that the Nation is falling behind in some international race and we must have higher achievement in order to be number one. Or, alternatively, as some argue, our standard of living is at risk and we need a more scientifically literate society to address that problem.

While such views may be laudable in many eyes, it begs a second educational function: The development of the individual. Throughout our history, education has been the principle tool of freedom—freedom of the individual to think what he or she chooses and to set whatever goals he or she aspires to, so long as those goals are not dysfunctional to the common good.

I believe the establishment of a single test or set of national goals for students would, ultimately, deprive us not only of the diversity on which the Nation was founded, and which sets us apart from all other developed nations, but also decrease the probability of greatness, especially among those individuals in our society who have the least opportunities.

Don't misunderstand my point. I'm not against high standards in literacy, numeracy or scientific knowledge. Public policy makers, such as governors, have a right to know how their investment in education is paying off in terms of the fundamentals of human capital. But I believe they should gain that information without unduly constraining the rights of the individual to set his or her own goals.

We have the technology to develop assessment procedures that can help us track our progress toward national goals without, at the same time, setting educational goals for individuals. That technology is already in place in the form of sampling procedures used by the National Assessment of Educational Progress, which, I might add, are being expanded to make the data meaningful at the state level. Whether an expanded—even criterion referenced—form of NAEP is the best vehicle for such a purpose is beyond the scope of my testimony.

The point is simply that we must avoid turning our need for better data about educational progress into a goal setting activity for individuals-which is exactly what will happen if the test becomes an every-student test. Such a test will quickly produce a national curriculum.

We must avoid confusing the need for measuring national progress against a desirable standard and using the standard to set individual goals.

Let me personalize my point. I have a daughter who is currently a Rotary Exchange student in Austria. Her life goal is to serve the Nation in an elected office and to teach political science. My wife and I have encouraged her and our other two daughters—one of whom is living in Argentina to perfect a foreign language and experience another culture to strive to become the best they can be and to do what truly inspires them. I would not have wanted their high school experiences dominated by a curriculum driven by a national goal of creating more scientists or mathematicians. Certainly I would want all students, including my daughters, to have skills and knowledge in these areas. But 1 also want them to have the freedom to pursue a course in the arts, in history, and even auto mechanics, if those were their personal goals.

I will not argue the point that we need more scientists and mathematicians. But the question is, is the best policy tool we have for achieving that goal a national test? Not from my experience.

Nor is a national test the best way to drive school reform and instruction. Remember that the challenge we face is less one of defining outcomes and more one of figuring out how to reach more students with a more powerful form of instruction and a better curriculum. We recognize the need for higher levels of learning, yet we must figure out how to get there.

The design of today's schools is rooted in the 19th Century. Today's instruction is based on models of lecture, drill and practice. The model is passive and assumes the child is an empty vessel in which to pump information.

Such a model served the emerging industrial society of the 1830's when it was important to teach immigrant children the importance of doing what they were told; it works less well in a society that requires creativity and problem solving. The passive character of the 1830's instructional model was reinforced by the textbook, which was invented to insure that students were thoroughly indoctrinated with values of the Protestant ethic—hardly the appropriate tool of an information society.

Today's curriculum was developed in the 1890's. It models the reductionist views of Newtonian science—a science being radically changed by physicists and other scientists into a more relativistic, systemic orientation to knowledge. To develop a national test on the current curriculum would freeze in place a system that is already out of date. And it would have a chilling effect on badly needed innovation.

Consider, for example, the changes we need to bring instructional practice into line with current views of psychology. From a psychological perspective, we know that individuals are not motivated by externally determined goals. Humans are goal driven, but those goals must be internally set and congruent with our own vision of ourselves.

We won't get more math or science pumped into kids' heads by willing it on them. It will only come about when our schools see their primary function as the empowerment of the individual to set high standards for him or herself, when we have redesigned instruction that is responsive to individual differences and inspires each student to strive for excellence, and when we have a new curriculum redesigned around more cognitively complex, systemic, contextually situated issues that are relevant to the 21st Century—issues such as the environment, energy, technology, productivity, space, social justice, or communications—issues that are not now central to the existing curriculum. A national test at this time would almost certainly freeze out the very kind of reform we need the most.

Some will argue the opposite: That we can use a national test to drive curriculum reform. While that might be possible in the abstract, I doubt that marketplace forces would permit it.

For example, even today, it is likely that a politically controlled, bureaucratic process would ask students about Gleick's theories of Chaos, fractal geometry or so-

ciobiology. The latter concept of sociobiology has been around for 20 years and it still isn't in any high school biology textbook I can find. Nonetheless, it is an accepted theory on the campuses of most major research universities.

These are the ideas we must challenge students with if we are really to be first in science and mathematics. But if the design and maintenance of a national test is left to a bureaucratic process, as it inevitably would be, the items on the test would not reflect what we should teach, but what we do teach, thereby freezing the past in place.

Let me put it another way. In the field of physics alone knowledge has changed so rapidly in the past decade that any effort we made to drive curricula through a test would be out of date before we could institutionalize the test and report the results.

If I were to put this in the context of systems science, I would say that the genius of America is that we have had the requisite variety to insure a vital, healthy Nation based on the ingenuity and creativity inherent in our diversity. To set a national standard for excellence, no matter what it was, would be to reduce that requisite variety and, ultimately to reduce the ingenuity and creativity that is our Nation's real strength.

As the director of a regional laboratory I have personally been working with the governors of two states that are seeking to restructure their state systems of education. They are as concerned as anyone could be that their states are competitive and their educational systems serve the larger social and economic goals of the State. But they are also sensitive to the needs of individuals and see their primary policy tool as inspiring and supporting dramatic change in education. Although I cannot and would not speak for them, I would be skeptical if they thought that a primary means for achieving that goal would be a national, every-pupil test.

Let me illustrate my point from the perspective of another state in which my laboratory works. We have a number of Indian Reservations in our region and students in schools on these reservations are as needy as any in the entire Nation. They are torn by conflicts between loyalty to their culture and the realities of making a living in a white-man's world.

In one situation in which we have worked, educators are trying to use the Indian culture to reconceptualize their challenge. For example, they are using concepts such as "the entrepreneur as warrior." And, they have defined as one kind of "bravery" what a young Native American does when he chooses to work outside of the reservation. In this school the science program is very holistic, focusing on the environment in which these students live, understanding the ecology and balance of nature needed to protect that environment.

I am skeptical of how well these students might do on a national science test and yet I think they are learning better science than students in many of the suburban schools in my region that teach science using classic laboratory methods. It would be very sad, indeed, if the spirit of change and vision that the educators in that school have were compromised by the necessity of scoring well on a test developed in Washington or even in the capital of the State.

In short, I am not testifying against the need to improve our Nation's performance in science and mathematics or any other content area. I am not testifying against the desirability of setting national goals. Nor am I testifying against the development of a single national mechanism for assessing our progress toward those goals. But I am testifying against the stultifying and counter-productive effect that a national every-pupil test would have.

Instead, I hope the Nation's attention can be refocused on the development of new curriculum and instructional procedures that will not only respect, but capitalize on the individual differences that have made this country unique in the history of civilization.

My specific recommendations are these:

1. Avoid a single high school graduation test applied to everyone. Use a sampling process to gather the data governors and others need to mark progress toward national goals.

2. If a national every-pupil test must be developed for political reasons, then gear it toward students leaving the elementary grades, perhaps sixth or seventh grade. There is a lot more consensus about what needs to be taught in the elementary grades. Moreover, a national curriculum based around such a test would not be nearly so destructive to differing individual needs.

3. Spark a revolution in the redesign of curriculum and instruction. The current restructuring movement has focused its energy on management changes—choice, site-based management, collaborative decision-making, etc. Research suggests these changes have almost no probability of affecting teacher/student interactions or the

curriculum—the places where change must occur if we are, in fact, to meet the national goals.

You might do that by considering such things as:

a. Creating a national trust fund such as the one Senator Kerry recently proposed. It would be important for such a fund to focus on innovation in curriculum and instruction. You might be interested in knowing that the state of Wyoming, one of the states in my region, created a $50 million trust fund for that purpose last week.

b. Fund some large-scale demonstrations—at least one in every state—that give schools the chance to really start all over and build, in effect, an alternative system. Again, in my region, with the help of my laboratory, that is exactly what the legislature in the state of South Dakota has just done. They will be creating eight pilot sites starting as early as this summer; these sites will have almost total freedom from state regulation to redesign their curriculum and instructional systems.

c. At the same time, consider the request of states like Missouri, another one of our states, to be free of Federal regulations in the way they use Federal education dollars so they can redeploy them holistically for the purposes of reform.

d. And, finally, consider asking the Office of Educational Research and Improvement to free the laboratories from the restriction that they cannot develop curriculum. OERI needs to be the national source of energy for innovation and change. They can do that by revitalizing their existing institutional contractors and funding more unsolicited field-initiated work that is not designed in response to tightly written Request For Proposals or grants announcements.

4. One final recommendation. Testing has a status in this country that makes it somewhat invulnerable to lay criticism. It is viewed as an objective, neutral, scientific process for arriving at truths that all of us are supposed to accept on faith. And yet we know that testing has many weaknesses: It can and has been used to sort people, it is notoriously susceptible to cultural biases, and it is very value laden. And in that context, remember that in arguing for a national test you are running counter to all of the other values that the reform movement has thus far espoused: Choice, site-based management, and authenticity in curriculum.

If we are to have a national test, whoever instigates it must be accountable for its abuses. We strongly advise you to explore how the Federal Government might help build standards of accountability for the testing and assessment industry—both public and private—that would create more public understanding of how the assessment industry affects education both negatively and positively. For example, if the policymakers instituting high stakes testing that can profoundly affect the future prospects of our children can't guarantee there won't be abuses, then the test must be accompanied by a process for insuring that students who take it, or their parents, are given the right of informed consent.

In summary, I would prefer that you put an enlightened set of national performance goals out there as a challenge to the Nation, and stimulate the creativity and imagination of teachers, administrators, parents and students to create new designs for education.

I close with an emphasis on students. My lab has been working with a network of schools who are literally rewriting the curriculum and using students to do it. You'd be surprised at what they can do. Challenge today's students and you will see real energy and real skill in figuring out how to make the system work better. Students know the problems their generation faces. We need to give them a chance to learn how to deal with them. I'm not sure we're wise enough to devise a test that they will find useful in the 21st Century.

Senator PELL. A question was raised earlier about the whether students could achieve more. I have done a simple test of my own, in talking with different high school groups in my State, and I ask them, "How many of you youngsters feel that you could learn more if you were more challenged? How many of you feel that you're working up to your capacity at this time?" Usually the ratio is about 9-to-1, the kids feel if they are more challenged they could learn more. And I don't how best to make them more challenged, whether it's through better teaching, whether it's through eliminating the 180 days per year of vacation, whether it's longer hours. That remains to be seen. But the kids themselves really feel that, if challenged, they could learn for more. It's an interesting experi-

ment I commend to my colleagues and others, when you're with a high school class, try it out.

I would like to ask one question to Mr. Shanker.

That is that the President's education policy advisory committee on which you serve, recommended to the President that he initiate the development of a national test. How did you reach this remarkable consensus, and how did you get there?

Mr. SHANKER. Well, with great difficulty. Well, it was raised some months ago, actually it was first raised last summer. And it was raised around the issue of if we have national goals, which we now do, how are we ever going to know around the year 2000 whether we've reached them or not; what do you mean by being first in the world in science and math; what do mean by being able to function well in these other areas?

So it became quite clear that the goals would be meaningless unless the goals were attached to some definition of standards and to some system of assessment.

We also discussed the question that was raised here about so many Americans thinking that education in general is in trouble but that their schools are okay, that we don't have a really good system of reporting to parents, to kids, and to individuals schools. Standardized tests which just tell you if you're above or below average or a certain percentile don't tell you what you're able to do, they merely tell you where you are in comparison.

At any rate, this was a very controversial subject. There were special meetings. There were experts. There were documents. There were papers traded back and forth. And eventually this became—I believe you have it—a statement of that group, which calls for support of the establishment of more than one, the notion that one would raise the question of a single curriculum.

There was a statement that was developed after a good deal of controversy and negotiation, and it does have the support of the overwhelming majority of the group.

I think, by the way, that that represents the same kind of change of opinion. If you had taken that vote when it was first proposed, I think there would have been an overwhelming rejection by that group. That came about as a result of a process of wrestling with, "Well, what are the alternatives?"

And, you know, we have had a long period of time without any national tests or system of tests or curriculum, and how well are we doing. And all these things, all these dangers that we talk about are German kids not able to think well? Do they not read? Or the Scandinavians? All these things we're talking about, all these terrible effects should occur in these other countries. Well, they don't.

Senator PELL. Well, thank you. I agree with you that historically we have 16,000 separate school committees each setting its own curriculum does boggle the mind, and in view of this system it is remarkable that we have done as well as we have.

Because of the exigencies of time, I will not ask any more questions at this time.

I turn to the ranking member.

Senator KASSEBAUM. Thank you, Mr. Chairman.

Mr. Shanker, I am interested in your comment on the consequences of not passing the test. If we take, say, for instance, Germany or Japan or Great Britain, it seems that what has driven educational quality there is more a respect for learning that is passed through the generations there that I think to a certain extent we have lost here. It has to come from students and parents, and there is a lot of concern.

Is it really rooted in a respect for learning that has driven, one could argue, Germany and Japan into sort of an elitist situation where, if you don't pass those tests, there are severe consequences?

In this country, basically, it seems to me we have wanted to be more all-inclusive. I think we should be—but, if we are going to have our standards raised and if we are going to do it by means of encouraging these kinds of tests, what troubles me are the students that don't make it. We are going to have to be prepared to provide the resources to have one-on-one teaching and to give greater emphasis to teaching parents how to be parents, how to work with their children in an educational forum.

I think that—and to a certain extent this was touched on in the other panel—if we don't provide the tools to help those who can't benefit from this test and those who benefit are ones who are going to anyway whether the tests are there or not—then it seems to me we have missed what we are trying to achieve here and it will only drive some students further down.

This goes back to why test? Educate America, I think, advocates just testing the 12th grade. Isn't that too late?

Mr. DiPATRI. Senator Kassebaum, our proposal calls for testing in 12th grade first. I think the President's panel suggested in grades 4, 8, and 12. Dr. Resnick earlier said at critical points, whether they're at grades 4, 8, and 12.

We are not suggesting that not be done, but the first step, the first priority should be to establish the outcomes that we want every high school student, regardless of where they live, regardless of ethnic background, what is it we want them to know when they graduate from high school, when they leave the 12th grade.

And so our proposal then says test in 12th grade, focus the outcomes at that level. What has happened in this country over the last 10 years in the minimum competency movement, 20 States have established graduation tests. What has happened is those tests are minimum skills, and the result has been a negative effect upon education.

And I would agree with my colleagues when they talk about simple multiple-choice tests, and it's because we have pegged the level at the 9th grade. Most States that administer graduation tests today test in either 9th or 10th grade, and, in effect, those skills are at the 7th or 8th grade level, so that students have multiple opportunities to pass. That is not what Educate America's proposal is talking about.

Senator KASSEBAUM. Mr. Shanker.

Mr. SHANKER. Well, I would like to respond this way, that I think it's wrong to describe the systems in these other countries as just an elitist system. The fact is that they have got, in many ways, a much more democratic system than we have. They don't give up on a kid if he doesn't learn something when he's 17, 18, or 19;

they've got educational institutions that continue to follow those people through the rest of their lives. And they attach various social consequences to not learning at various times, so that they don't end up with a huge underclass of people who constantly have to be supported, because they don't get them.

Before World War II, both Germany and the United States had about the same percentage of high school graduates and about the same percentage of youngsters going to college. After the war, both countries decided they wanted almost all kids or they wanted all kids to graduate high school and they wanted many more to go college, but each country had a different philosophy:

The Germans essentially said we want more kids in college; therefore, educate more kids to be able to do college work. And they lifted the level of millions of kids.

In the United States, we said we want more kids to go to college; these kids won't be able to learn enough to go to college, so let's lower our standards.

Now, I think what has been shown in these other countries is that many more kids can achieve, and our kids can too.

Now, I think that this business of attaching consequences, there is a lead piece in Commentary magazine this month by Barbara Lerner, which deals with the minimum competency tests that were introduced in the 1970's. A lot of people said that if you introduce minimum competency test, lots of kids are going to drop out because they're going to look at them and they're going to say, "I am never going to be able to pass that, I won't be able to get my high school certificate, I might as well leave now."

Well, the fact is that the dropout rate has not increased and the fact is that black and Hispanic youngsters, about 80 to 90 percent of them, on NAEP used to be down in the illiterate or semiliterate category. There is almost nobody in that category now. In other words, by adopting standards, by telling youngsters there is a consequence attached, by giving teachers the flexibility to figure out new ways of reaching those standards, we have brought huge numbers of kids up from those bottom levels.

And at the same time that we brought them up from the bottom levels, it's not surprising that the SAT scores of black students have gone up. Why? Because if you've got more people who can read at decent levels, who are moved up from illiteracy, you're going to have a greater pool of youngsters who then start reading things that are worthwhile. So that are concentrating on minimums doesn't mean that that is where you have to stop. I think that we need a closer look at this.

Now, we do have a problem in some areas, like what do we teach, do we have agreement on history or on literature? But I think there is agreement in the field of mathematics. The math teachers and professionals move ahead on that. I think that we do have the ability to test reading and writing, and we ought to do that. Those are areas of agreement.

Yes, it's going to cost money. No, it should not be done through multiple-choice. Yes, we shouldn't wait till the kids are 17 or 18. But let's move ahead in areas where there is agreement and then let's put the pressure on the other groups to come up with agree-

ment in curriculum areas so that we can move ahead in the other areas as well.

Senator KASSEBAUM. I certainly agree with much of what you're saying, Mr. Shanker, I would only say when I spoke about elitism, it was more going back to a respect for learning, which has absolutely propelled the Germans to make sure that those standards were high, and to the dedication of the families and the teaching society to make sure that this was going to be accomplished. It was a focus and a respect that, on the whole, has not had a priority here in this country recently. It's hard to get our best and our brightest into teaching today.

Mr. SHANKER. Well, but part of this has to do with the fact that don't you lose respect for learning when basically 80 to 90 percent of the youngsters who are in colleges and universities in this country are getting their junior high school and high school education and calling it a college education?

Senator KASSEBAUM. I couldn't agree with you more.

Mr. SHANKER. I mean, I think that when you have no standards and when you're engaged in a sort of great inflation, essentially that lowers the value of education.

Senator KASSEBAUM. I certainly agree.

Thank you, Mr. Chairman.

Senator PELL. Thank you, Senator.

Senator Bingaman.

Senator BINGAMAN. Well, thank you, Mr. Chairman.

Dr. DiPatri, let me as you what your group, Educate America, estimated the cost of the test as being; what you recommended?

Mr. DIPATRI. $30 per student, and roughly $90 million for Federal appropriation. But that is based on what we believe to be a NAEP-like test. I said it earlier, and I said often it's misrepresented and somehow has been today. We believe a combination of multiple-choice and performance measures. I wouldn't throw out multiple-choice tests because they are characterized as simple tests. A multiple-choice test can be difficult. Doctors in our States take tests, the majority of which are multiple-choice. The advanced placement is a multiple-choice test.

Senator BINGAMAN. But the $30 per test figure is what you would estimate to do the good test that you would intend to have administered; is that right?

Mr. DIPATRI. That's correct.

Senator BINGAMAN. OK. And who do you recommend administer the test or prepare the test and then administer it? How do you propose to implement your recommendation?

Mr. DIPATRI. Well, it would be very similar to what earlier testimony was given, that a national board be established, not unlike the National Assessment Governing Board, charged with the responsibility with that appropriation to issue contracts to the major testing firms in this country who do it now, for up to 1.6 million students per year are administered the SAT, that these national testing firms would bid competitively and would bid on tests, as I described them, that had a balance of multiple-choice and performance. And it's happening in places today. NAEP is the perfect example.

Senator BINGAMAN. And the recommendation is that the Congress pass a law requiring 12th graders to take this test in November but not requiring that anybody pass it in order to graduate; is that correct?

Mr. DIPATRI. That's correct. But it's important that we tie in the consequences that Al Shanker mentioned. We think that business and industry, if the business and industry could establish a system just as college admissions counselors do today, they don't just ask the student for SAT scores, they ask for a total record, "Let me see your transcript, let me see your class rank, let me see your SAT's." And that then supplements the transcript. This objective measure, zero to 200 in each of the six areas.

So if I said to this room today, "I have a student in 12th grade and he has a 740 in math on his SAT and a 380 in verbal," everybody here knows what I just said. You could picture that student, and you know that student is not going to be an engineer; he is probably a journalism student or a very bright, verbal young man, but may have difficulty in math. The same could happen in these six areas.

Senator BINGAMAN. I think you got that backwards. I think you said 740 in math.

Mr. DIPATRI. Excuse me. I'm sorry. Then I got that backwards.

Mr. SHANKER. Erase that multiple-choice.

Mr. DIPATRI. Thanks, Al.

Senator BINGAMAN. Mr. Shanker, let me ask you if your group that advised the President that we should have national tests, did you also advise him as to where the resources would come from to administer the test, to prepare and administer the tests and how that would be done, if it would be done through the Department of Education or how that would be accomplished?

Mr. SHANKER. I don't think that the view was that this would be—I think the statement is that the President should support the establishment of more than one system. And I was not at the last meeting, so I am going to have to stay with the statement. I have had some telephone calls with people there.

But my view is that a system which would have, which would be more like what Lauren Resnick and Marc Tucker are talking about, which essentially is to develop a system through voluntary efforts, which would provide such a good system that there would be a substantial amount of political pressure to buy into it.

As I think you need support from the top in terms of the idea and its credibility and its legitimacy. But I also think that you need some ability for some experimentation and some variety. We aren't coming at it the way the Europeans do with one that started at the top, and it's not going to be accepted immediately.

I think that if you had two or three systems where different States got on board and supported them and perhaps at some point with some support, if a certain number of States got into it and the subject matter and professional groups, I think what is viewed now is not a single system from the top but the encouragement of several systems which would have to gain their legitimacy because people would take a look and start asking why isn't our school system requiring this or why isn't our State doing this; that it's viewed as putting something out there which is so good that there

is a lot of political pressure to do it rather than having Congress say every kid has to do it.

Senator BINGAMAN. So you would not recommend that we mandate that testing be done at 4th, 8th, and 12th grades in all schools in the country against some common standard?

Mr. SHANKER. I would not. I would keep NAEP for what it is. I would give it some more money so it could do some better and more imaginative and more performance-oriented things and not abandon but move away from some of the multiple-choice. I would keep that as a very valuable indicator.

Then I would give assistance to the voluntary establishment of a number of other systems of examination which would be designed to do several things. One would be to measure progress toward national goals, but the other one and in the report "America's Choice," I think the notion of creating a certificate for kids around the age of 16 which would be tied to the incentive of no after-school or part-time employment until you get this, which would be designed to get everyone to achieve at least a certain minimum standard, would be very important.

Mr. HUTCHINS. Senator, might I comment?

Senator BINGAMAN. Yes, certainly.

Mr. HUTCHINS. We are not opposed to setting standards, but opposite, and feel that the process of setting standards could be as important as the standards themselves. You made a comment in your question that suggested the creation of a single national panel. I think our view would be that there needs to be a lot of grass-roots, bottoms-up participation in that process. We would hope the process might look like something having State summits operating under some kind of framework, the kind of which Dr. Resnick suggested, in which the States contribute their suggestions as to what those standards would be, and that that is rolled into a national process.

But to have a single panel at the national level that sets those standards we think would be counterproductive.

Senator BINGAMAN. Thank you very much, Mr. Chairman.

Senator PELL. Thank you very much.

Without objection, we ought to insert in the record, I think, the "Monitoring and Achieving the National Education Goals," the document that Mr. Shanker put out.

[The document referred to follows:]

The Basic Premise

Each of the six national goals requires a form of measurement so that progress or problems in achieving them can be assessed accurately. Goal 3 calls for each student to have mastered challenging subject matter by completion of grades 4, 8 and 12. Even if "challenging subject matter" were clearly defined, there is no current measurement system upon which the society can rely to determine whether this goal is achieved for each child, although the National Assessment of Educational Progress (NAEP) is a start on measuring it for the nation as a whole.

This situation reflects a fundamental problem in the current education system: There are no tests based on absolute standards of content and proficiency by which each student, parents, teachers, school administrators, any level of government or the general public can judge student performance, compare that performance to other students, or aggregate results to make comparisons to other classes, schools, states or countries.

- There are many tests for individuals available commercially or in a number of states, but they are not comparable to each other and none is derived from a national consensus on what children of a given age should know.
- Without such tests, the host of education reform strategies cannot be targeted effectively or judged properly on the basis of their results. Accountability systems remain judgmental, not objectively based.
- Sample surveys, like NAEP, are necessary for assessing the national (and eventually, perhaps, state) educational status, but are not helpful for understanding individual performance and diagnosing educational improvement needs at the individual, class, school or school district levels.

While public leaders and the public generally have both long accepted the fact that the educational system fails to meet the needs of the nation, few school systems and parents believe that their own children are failing to receive adequate preparation. Real change in the system requires knowing who is succeeding and who is not and basing reforms on that knowledge.

The Ideal

A. Standards
   1. Standards for knowledge of skills and subjects to be mastered should be established for various ages or grade levels beginning no later than age 9 or grade 4. These should be developed with maximum consensus among experts in each field and others and should reflect the highest levels of knowledge achieved at various ages worldwide. They should describe both content and level of difficulty.

2. Standards should be developed for English, Math, Science, History and Geography, and for reasoning and problem solving skills for increasing age levels.

3. Teachers should be informed of the standards and would be expected to teach to the standards.

B. Tests

1. At the key transition stages (however expressed: age 9, 13, 17, or 4th, 8th and 12th grade), a test or tests should be available to be given to every student, regardless of income level, disability, or English proficiency. The results of these tests will tell that student, the student's parents, teachers and others, whether the student has achieved minimum competency for his or her age, is proficient or excelling. The test should provide enough information to identify remediation needs for students and indicate system (class, school, LEA, state) improvement requirements.

2. The tests should cover the subject matter appropriate to each age, as established in the standards. The test should focus on the core subjects (e.g., at the youngest level, reading, computation, writing) but cover others that are desirable as well.

3. The tests should combine (with increasing complexity as age increases) assessments of mastery of subject matter and acquisition of generic (reasoning, problem solving) skills.

4. The tests should be devised with maximum consensus among experts in each subject field and the testing field, and others.

Implementation

While the technology for a useful assessment program is available, developing the standards and tests will take time and money. The first priority should be for 9 year-olds (4th grade). These children are at a point in their education when they should have achieved the basic skills necessary for learning increasingly more complex subject matter and developing increasingly more sophisticated skills. Uncorrected failure at this stage becomes more difficult to cure as the child ages.

The President, working with the Governors, should:

1. Facilitate the definition of the new higher world standards using experts in the appropriate subject fields and others.

2. Cause to be developed at least two tests which reflect these new standards. Existing tests appropriately modified to measure the new standards and yield comparable data might also be appropriate. These tests would be available to be given to every child everywhere in the U.S. (but the Federal government would not require anyone to take them.)

3. Create and issue a regular report on the tests given by states so that school officials and parents know how well the various tests reflect the new standards.

The new standards and assessments should be developed in such a way that the content and methodology has the support of the overwhelming majority of experts. Once devised, a campaign by business, other interest groups and the public sector in support of using these standards should be mounted. Parents and others should be able to challenge States and school districts on what they are using to measure student progress that is as good or better than the uniform tests.

In addition, the President should:

1. Establish as his public policy commitment the rapid development and use of such standards and tests as one of the highest national priorities for catalyzing education reform.

2. Support the National Education Goals Panel's effort to develop a useful assessment program. He should encourage the panel to establish immediately a process for involving the Chief State School Officers, educators, the business community, experts in subject matter, generic skills and test methodology in the development of the standards and tests. An ambitious time table should be publicly announced. The 4th grade/9 year-olds program should be ready for use within two years.

Senator PELL. I thank all of you for being with us.

We now turn to the third and final panel: Marc Tucker, the president of the National Center on Education and the Economy; Maureen Daniels, representing the National Education Association; Burton Faldet, president of Test Consultants, Association of American Publishers, who is accompanied by Michael Melody.

Thank you all for being with us.

We will start off with Mr. Tucker.

STATEMENTS OF MARC S. TUCKER, PRESIDENT, NATIONAL CENTER ON EDUCATION AND THE ECONOMY, ROCHESTER, NY; MAUREEN DANIELS, SIXTH-GRADE TEACHER, LYNBROOK ELEMENTARY SCHOOL, FAIRFAX, VA, ON BEHALF OF NATIONAL EDUCATION ASSOCIATION; AND BURTON W. FALDET, PRESIDENT, TEST CONSULTANTS, LTD., ASSOCIATION OF AMERICAN PUBLISHERS, WASHINGTON, DC, ACCOMPANIED BY MICHAEL E. MELODY, SENIOR VICE PRESIDENT, COLLEGE AND TEST PUBLISHING, HOUGHTON MIFFLIN COMPANY

Mr. TUCKER. Thank you, Mr. Chairman. I am Marc Tucker, director and president of the National Center on Education and the Economy, of Rochester, NY.

Mention was made just a moment ago of a report that we issued last year, "America's Choice: High Skills or Low Wages." In the course of preparing that report, we visited seven nations, among which was the United States, asking ourselves of those other nations why it is that they are able to achieve so much higher success in education and skill-building than we.

In every one of those nations that we visited—Singapore, Japan, England, Ireland, Sweden, and Germany—they had strong standards. Those standards were either very high curriculum standards or they were set by examinations or both. I was stunned when thinking about it, in contrast to those nations, to realize that in this country only the students who wish to go and expect to go to selective colleges have any incentive to get more than a high school diploma in the United States. And what you need to get a high school diploma in the United States is to show up about two-thirds of the time and not cause too much trouble. It has nothing to do with what you know and can do. And for that reason, there is virtually no connection, for the vast majority of kids in the United States, between what they do in school and what they want for themselves later in life.

That is not true in the countries with which we compete. As you heard a moment ago, we proposed in that report that this Nation create a certificate of initial mastery, that it represent a standard, benchmarked to the highest world standard for kids of 16, and that we award kids a certificate of initial mastery when they meet that standard.

Now, I should make it clear what is meant here. What we are talking about is not having a scaled exam. And this, by the way, is in great contrast to almost every other proposal you have heard about this morning. We are not talking about a scaled exam which simply says, "With respect to these goals, here is where you fall: At the top, in the middle, or on the bottom."

What we are talking about is a certificate that says, "You have achieved a world-class level of mastery or you haven't." That's it. And that benchmarked standard would be the gateway to what kids want out of life. Those are the consequences to which Al Shanker just referred.

Now, we joined forces last fall with a learning research and development center, co-chaired by Lauren Resnick, from whom you heard two panels ago, and we are now busy putting together a large project, the purpose of which is to create an examination system of the kind that she alluded to.

We have about $2.5 million of private foundation money. Some 20 States and cities are joining us as our partners in this effort.

I brought with me an overview of the description of that project, which I would, with your permission, like to enter into the record.

Senator PELL. Without objection.

[The document referred to follows:]

92

America's Choice: high skills or low wages!

The Report of
The Commission on the Skills of the American Workforce

June 1990

NATIONAL
CENTER
ON
EDUCATION
AND THE
ECONOMY

97

EXECUTIVE SUMMARY

The Problem

Since 1969, real average weekly earnings in the United States have fallen by more than 12 percent. This burden has been shared unequally. The incomes of our top 30 percent of earners increased while those of the other 70 percent spiraled downward.

In many families, it now takes two people working to make ends meet, where one was sufficient in the past.

The United States is in the midst of the second longest economic expansion in its history. But that expansion is built largely on the fact that 50 percent of our population is employed compared with 40 percent in 1973. Forty million new jobs were created as the 'baby boom' generation reached working age, and more women entered the workforce. More of us have been working so we produced more.

However, workforce growth will slow dramatically in the 1990's. We can no longer grow substantially just by adding new workers.

The key to maintaining, to say nothing of improving, our standard of living is productivity growth — more products and services from every member of the workforce.

But during the past two decades, our productivity growth has slowed to a crawl. It now takes nearly three years to achieve the same productivity improvement we used to achieve in one year.

If productivity continues to falter, we can expect one of two futures. Either the top 30 percent of our population will grow wealthier while the bottom 70 percent becomes progressively poorer or we all slide into relative poverty together.

The Task

To ensure a more prosperous future, we must improve productivity and our competitive position. We cannot simply do this by using better machinery, because low wage countries can now use the same machines and can still sell their products more cheaply than we can.

The key to productivity improvement for a high wage nation lies in the third industrial revolution now taking place in the world. The steam engine and electric motor drove the first two industrial revolutions, causing profound changes in work organization. This boosted productivity, quality and living standards dramatically. The creation of the modern factory in the 1800's and mass production in the 1900's followed these technology breakthroughs.

42-921 0 - 91 - 4

The advent of the computer, high speed communication and universal education are heralding a third industrial revolution, a revolution the key feature of which is high performance work organization.

The Organization Of Work In America

The organization of America's workplaces today is largely modeled after the system of mass manufacture pioneered during the early 1900's. The premise is simple: Break complex jobs into a myriad of simple rote tasks, which the worker then repeats with machine-like efficiency.

The system is managed by a small group of educated planners and supervisors who do the thinking for the organization. They plan strategy, implement changes, motivate the workers and solve problems. Extensive administrative procedures allow managers to keep control of a large number of workers. This form of work organization is often referred to as the 'Taylor' model.

Most employees under this model need not be educated. It is far more important that they be reliable, steady and willing to follow directions.

But in the world's best companies, new high performance work organizations are replacing this 'Taylor' method. These companies are using a new approach to unleash major advances in productivity, quality, variety and speed of new product introductions.

Mass production methods will continue to produce high volume, inexpensive goods and services for a long time to come. But what the world is prepared to pay high prices and high wages for now is quality, variety and responsiveness to changing consumer tastes, the very qualities that the new methods of organizing work make possible.

'Tayloristic' methods are not well suited to these goals. Firms struggling to apply the traditional methods of work organization to more complex technologies, more frequent product introductions, increased quality requirements and proliferating product variety often create cumbersome and inefficient bureaucracies.

The new high performance forms of work organization operate very differently. Rather than increasing bureaucracy, they reduce it by giving front-line workers more responsibility. Workers are asked to use judgment and make decisions. Management layers disappear as front-line workers assume responsibility for many of the tasks — from quality control to production scheduling — that others used to do.

Work organizations like these require large investments in training. Workers' pay levels often rise to reflect their greater qualifications and responsibilities. But the productivity and quality gains more than offset the costs to the company of higher wages and skills development.

Despite these advantages, 95 percent of American companies still cling to old forms of work organization.

Is There A Skills Shortage In The United States?

Because most American employers organize work in a way that does not require high skills, they report no shortage of people who have such skills and foresee no such shortage. With some exceptions, the education and skill levels of American workers roughly match the demands of their jobs.

Our research did reveal a wide range of concerns covered under the blanket term of 'skills.' While businesses everywhere complained about the quality of their applicants, few talked about the kinds of skills acquired in school. The primary concern of more than 80 percent of employers was finding workers with a good work ethic and appropriate social behavior: 'reliable,' 'a good attitude,' 'a pleasant appearance,' 'a good personality.'

Most employers we interviewed do not expect their skill requirements to change. Despite the widespread presumption that advancing technology and the evolving service economy will create jobs demanding higher skills, only five percent of employers were concerned about a skills shortage. These were mainly large manufacturers, financial service organizations and communications companies.

The reason we have no skills shortage today is that we are using a turn-of-the-century work organization. If we want to compete more effectively in the global economy, we will have to move to a high productivity work organization.

How We Prepare Our Front-Line Workers For Work

More than 70 percent of the jobs in America will not require a college education by the year 2000. These jobs are the backbone of our economy, and the productivity of workers in these jobs will make or break our economic future.

No nation has produced a highly qualified technical workforce without first providing its workers with a strong general education. But our children rank at the bottom on most international tests — behind children in Europe and East Asia, even behind children in some newly industrialized countries.

More than any other country in the world, the United States believes that natural ability, rather than effort, explains achievement. The tragedy is that we communicate

to millions of students every year, especially to low-income and minority students, that we do not believe that they have what it takes to learn. They then live up to our expectations, despite the evidence that they can meet very high performance standards under the right conditions.

Unlike virtually all of our leading competitors, we have no national system capable of setting high academic standards for the non-college bound or of assessing their achievement against those standards.

America may have the worst school-to-work transition system of any advanced industrial country. Students who know few adults to help them get their first job are left to sink or swim.

Only eight percent of our front-line workers receive any formal training once on the job, and this is usually limited to orientation for new hires or short courses on team building or safety.

The American post-secondary education and training system was never designed to meet the needs of our front-line workers. The system is a combination of education programs for full-time college students and short term training for the severely disadvantaged, and can be difficult to access. Because employers have not set training standards, few students can be sure that there is a market for the courses they pursue. Education is rarely connected to training and both are rarely connected to an effective job service function.

Another Way

While the foreign nations we studied differ in economy and culture, they share an approach to the education and training of their workers and to high productivity work organization.

- They insist that virtually all of their students reach a high educational standard. We do not.
- They provide 'professionalized' education to non-college bound students to prepare them for their trades and to ease their school-to-work transition. We do not.
- They operate comprehensive labor market systems which combine training, labor market information, job search and income maintenance for the unemployed. We do not.
- They support company based training through general revenue or payroll tax based financing schemes. We do not.
- They have national consensus on the importance of moving to high productivity forms of work organization and building high wage economies. We do not.

Our approaches have served us well in the past. They will not serve us well in the future.

The Choice

Americans are unwittingly making a choice. It is a choice that most of us would probably not make were we aware of its consequences. Yet every day, that choice is becoming more difficult to reverse. It is a choice which undermines the American dream of economic opportunity for all. It is a choice that will lead to an America where 30 percent of our people may do well — at least for awhile — but the other 70 percent will see their dreams slip away.

The choice that America faces is a choice between high skills and low wages. Gradually, silently, we are choosing low wages.

We still have time to make the other choice — one that will lead us to a more prosperous future. To make this choice, we must fundamentally change our approach to work and education.

1. Problem: Two factors stand in the way of producing a highly educated workforce: We lack a clear standard of achievement and few students are motivated to work hard in school. One reason that students going right to work after school have little motivation to study hard is that they see little or no relationship between how well they do in school and what kind of job they can get after school. Other advanced industrial nations have stringent performance standards that virtually all students must meet at about age 16 and that directly affect their employment prospects.

Recommendation: ***A new educational performance standard should be set for all students, to be met by age 16. This standard should be established nationally and benchmarked to the highest in the world.***

We propose that all American students meet a national standard of educational excellence by age 16, or soon thereafter. Students passing a series of performance based assessments that incorporate the standard would be awarded a Certificate of Initial Mastery.

Possession of the Certificate of Initial Mastery would qualify the student to choose among going to work, entering a college preparatory program or studying for a Technical and Professional Certificate, described below.

Creation of the Certificate of Initial Mastery standard would require a new approach to student performance assessment. We recommend the creation of new performance based examinations for which students can explicitly prepare. The assessment system would provide multiple opportunities for success rather than a single high stakes moment of possible failure. Most important, the examination, though set at a very high standard, is not intended as a sorting mechanism on the pattern of virtually all the major tests now in use. Our goal is to

BEST COPY AVAILABLE

set a tough standard that almost everyone will reach, although not all at the same time.

Once created, this system would establish objective standards for students and educators, motivate students and give employers an objective means to evaluate the accomplishments of students.

2. Problem: More than 20 percent of our students drop out of high school — almost 50 percent in many of our inner cities. These dropouts go on to make up more than one third of our front-line workforce. Turning our backs on those dropouts, as we do now, is tantamount to turning our backs on our future workforce.

Recommendation: ***The states should take responsibility for assuring that virtually all students achieve the Certificate of Initial Mastery. Through the new local Employment and Training Boards, states, with federal assistance, should create and fund alternative learning environments for those who cannot attain the Certificate of Initial Mastery in regular schools.***

All students should be guaranteed the educational attention necessary to attain the Certificate of Initial Mastery by age 16, or as soon as possible thereafter. Youth Centers should be established to enroll school dropouts and help them reach that standard. Federal, state and local funds should be raised or reallocated to finance these dropout recovery programs. Once the Youth Centers are created, children should not be permitted to work before the age of 18 unless they have attained the Certificate of Initial Mastery or are enrolled in a program to attain it.

3. Problem: Other industrial nations have multi-year career-oriented educational programs that prepare students to operate at a professional level in the workplace. Graduates of these programs have the skills to hit the ground running when they get their first full-time job at age 19 or 20. America prepares only a tiny fraction of its non-college bound students for work. As a result, most flounder in the labor market, moving from low paying job to low paying job until their mid-twenties, never being seriously trained.

Recommendation: ***A comprehensive system of Technical and Professional Certificates and associate's degrees should be created for the majority of our students and adult workers who do not pursue a baccalaureate degree.***

Technical and Professional Certificates would be offered across the entire range of service and manufacturing occupations. A student could earn the entry-level occupation specific certificate after completing a two- to four-year program of combined work and

study, depending upon the field. A sequence of advanced certificates, attesting to mastery of more complex skills, would be available and could be obtained throughout one's career.

The Secretary of Labor should convene national committees of business, labor, education and public representatives to define certification standards for two- to four-year programs of professional preparation in a broad range of occupations. These programs should combine general education with specific occupational skills and should include a significant work component.

Students could pursue these programs at a wide variety of institutions accredited to offer them, including high schools, community colleges and proprietary schools. The system should be designed to make it possible for students to move easily between the Certificate programs and college.

A means should be established to ensure that all students can receive financing to pursue these programs.

4. Problem: The vast majority of American employers are not moving to high performance work organizations, nor are they investing to train their non-managerial employees for these new work organizations. The movement to high performance work organizations is more widespread in other nations, and training of front-line workers, funded in part by national assessments on employers or general public revenues, is commonplace.

Recommendation: ***All employers should be given incentives and assistance to invest in the further education and training of their workers and to pursue high productivity forms of work organization.***

We propose a system whereby all employers will invest at least one percent of their payroll for the education and training of their workers. Those who do not wish to participate would contribute the one percent to a general training fund, to be used by states to upgrade worker skills. We further recommend that public technical assistance be provided to companies, particularly small businesses, to assist them in moving to higher performance work organizations.

5. Problem: The United States is not well organized to provide the highly skilled workers needed to support the emerging high performance work organizations. Public policy on worker training has been largely passive, except for the needs of a small portion of the severely disadvantaged population. The training system is fragmented with respect to policies, administration and service delivery.

Recommendation: ***A system of Employment and Training Boards should be established by Federal and state governments, together with local leadership, to organize and oversee the new school-to-work transition programs and training systems we propose.***

We envision a new, more comprehensive system where skills development and upgrading for the majority of our workers becomes a central aim of public policy.

The key to accomplishing these goals is finding a way to enable the leaders of our communities to take responsibility for building a comprehensive system that meets their needs. The local Employment and Training Boards for each major labor market would:

- Take responsibility for the school-to-work and Youth Center-to-work transition for young people.
- Manage and oversee the Youth Centers.
- Manage and oversee a 'second chance' system for adults seeking the Certificate of Initial Mastery.
- Manage and oversee the system for awarding Technical and Professional Certificates at the local level.
- Manage a labor market information system.
- Manage and oversee the job service.
- Coordinate existing programs.

The states would need to create a parallel structure to support the local Boards, coordinate statewide functions and establish state standards for their operation.

In Conclusion

America is headed toward an economic cliff. We will no longer be able to put a higher proportion of our people to work to generate economic growth. If basic changes are not made, real wages will continue to fall, especially for the majority who do not graduate from four-year colleges. The gap between economic 'haves' and 'have nots' will widen still further and social tensions will deepen.

Our recommendations provide an alternative for America. We do not pretend that this vision will be easily accepted or quickly implemented. But we also cannot pretend that the status quo is an option. It is no longer possible to be a high wage, low skill nation. We have choices to make:

- Do we continue to define educational success as 'time in the seat,' or choose a new system that focuses on the demonstrated achievement of high standards?
- Do we continue to provide little incentive for non-college bound students to study hard and take tough subjects, or choose a system that will reward real effort with better pay and better jobs?

- Do we continue to turn our backs on America's school dropouts, or choose to take responsibility for educating them?
- Do we continue to provide unskilled workers for unskilled jobs, or train skilled workers and give companies incentives to deploy them in high performance work organizations?
- Do we continue in most companies to limit training to a select handful of managers and professionals, or choose to provide training to front-line workers as well?
- Do we cling to a public employment and training system fragmented by institutional barriers, muddled by overlapping bureaucracies and operating at the margins of the labor market, or do we choose a unified system that addresses itself to a majority of workers?
- Do we continue to remain indifferent to the low wage path being chosen by many companies, or do we provide incentives for high productivity choices?

Taken together, the Commission's recommendations provide the framework for developing a high quality American education and training system, closely linked to high performance work organizations. The system we propose provides a uniquely American solution. Boldly executed, it has the potential not simply to put us on an equal footing with our competitors, but to allow us to leap ahead, to build the world's premier workforce. In so doing, we will create a formidable competitive advantage.

The status quo is not an option. The choice we have is to become a nation of high skills or one of low wages.

The choice is ours. It should be clear. It must be made.

Mr. TUCKER. I should tell you that what we are doing is inspired, in many ways, by what the Europeans have done, but it is not a copy. I just essentially gave you one of the most important respects in which that is true; that is to say we think it would be a terrible error to set up a national examination system in the United States the purpose of which is to do what the Europeans do and what almost every American testing system does, which is to sort out students.

This country more than almost any other that I know of in the world is a country that says when kids are in elementary school, this kid's got it and this kid doesn't. And because of the estimates that we make of their ability, very early in life, we assign to them different educational futures, work of different difficulty and different implicit standards.

Were we, in my judgment, to put in place in this country yet another sorting mechanism, the purpose of the test being to sort, we would simply make worse the toughest problem we already face, which is that we consign a third or more of our kids, before they have left elementary school, to failure.

This idea that we have is very different. It is an idea that says these kids can make it, they will make it, and the whole purpose of the system is to help them make it. By the way, what we are proposing is not a system which says you test at X age, everybody sits down at that age, and then we sort you out on performance. It is, rather, a system which says, "You have as long as it takes to get there, but you're going to get there."

Now let me just make a couple of points about the Congress. I think it would be a great mistake for the Congress at this point to legislate a test or to create a testing organization. I think it's not needed and not wise. I think in fact it would make it more difficult for us to get to an examination system. What you can do is help to create a mechanism through which we can agree on a framework. You can provide resources for research and development on assessment to many groups working on this issue at once, which I think is absolutely critical. And most critical of all is to provide resources through OERI for the restructuring of American education.

It would be an utter disaster, in my view, to have a testing system in which the whole burden of succeeding falls on the student. If it does not fall on the system, we will be worse off than when we started.

Senator PELL. Thank you very much indeed, Mr. Tucker.

[The prepared statement of Mr. Tucker follows:]

PREPARED STATEMENT OF MR. TUCKER

Thank you, Mr. Chairman, and the members of this committee, for the opportunity to testify today. I am president of the National Center on Education and the Economy, an organization dedicated to bringing the skills of the American workforce up to world class standards. John Sculley, CEO of Apple Computer, is our board chairman. Among our trustees are former governors James Hunt and Thomas Kean, Kodak CEO Kay Whitmore and former Labor Secretary Ray Marshall.

I can think of no idea offered in the last two decades that has had as much promise for American education—no policy proposal more worthy of Congress's attention—than the proposal for national examinations of student performance. This is not a new idea to this committee, or, especially, to you, Mr. Chairman, but it appears to be an idea whose time has come.

Two months ago, I was in Indonesia. I had occasion then to ask the president of the American Chamber of Commerce in Jakarta how the skills of the work-bound graduates of Indonesia's primary schools compare to those of America's high school graduates who enter the workforce after school. He replied that the skills of both groups of graduates are about the same, but, because Indonesian employers get the cream of the crop and the American employers get the bottom, Indonesian employers get more skilled workers than do the Americans. The prevailing wages for manufacturing jobs are far below what they are here. Little wonder that the Indonesian economy is burgeoning, attracting employers from all over the world, and that real wages are rising there while they are falling fast in the United States.

But, can it be that the average graduate of 12 years of schooling in the U.S. knows no more than the average graduate of the 6-year primary system on the island of Java? It sounds preposterous until one looks at the evidence. One American firm after another is reporting that when they administer tests of basic skills in reading and math to recent school graduates applying for entry level jobs, the applicants typically fail to reach the seventh grade level of achievement. For probably half of our students, we are producing 6 years of education in 12 years of seat time!

High school diplomas do not certify that high school graduates have twelfth grade skills. In many states, they certify only that the holder has spent 12 years in school. Others states do now have a basic skills requirement for high school graduates, but they are not set at the twelfth grade level of difficulty, or even remotely close to that.

Last June, our organization released the report of our commission on the Skills of the American Workforce, America's Choice: High Skills or Low Wages! The commission is chaired by Ira Magaziner, an international business strategy consultant, and co-chaired by Hillary Clinton, an Arkansas attorney, and Bill Brock and Ray Marshall, both former labor secretaries. The commission's report was based on the largest international study ever undertaken of workforce skill requirements and the measures that countries on three continents take to meet those requirements. The commission reaffirmed what many others concluded earlier—that the preparation of American front-line workers lags far behind that of our leading competitors. Recommendation number one of the commission was to create a national examination system in the United States and to give a Certificate of Initial Mastery to students who meet the standard set by the examination.

Why? Because the commission observed that every country that has higher education performance than ours has high standards for its students and enforces those standards. They either set a detailed national curriculum and make sure that curriculum is taught in every school in the country or they have national examinations, or both. For many reasons, this country is hardly likely to embrace an official national curriculum. The only alternative is national examinations.

Much of the power of examinations in other countries comes not from the examination itself, but rather from the way those countries use examinations to induce students to put real effort into their school work.

In the United States alone of all the countries we studied, the only students who have any incentive to take tough courses or to study hard are the students who plan to go to selective colleges, less than a quarter of all students. For all the others, only a high school diploma is required to go to college or get a job. All one needs to get a diploma is to show up with reasonable frequency and not cause trouble. Why take a real science course or algebra and risk a failing grade when you can take gut courses and be assured of getting the only credential you need?

But in the leading countries with which we compete, it is typically the case that the courses you take, the grades you get, the recommendations you have from your teachers and—especially—the scores you get on examinations, all combine to determine not only your eligibility for further education but also your access to good jobs, including those that do not require a college education. So students—almost all students—put a lot more effort into school than the vast majority of our students do. Clearly, countries that use examinations to provide a strong motivation to students to take tough courses and study hard will produce much better educated students than countries that do not.

The commission observed something else of great importance. In the United States, test security is paramount. The idea is to keep the items secret, and to create a test that is independent of the curriculum being taught. This means that one cannot study for the test. Many of our competitors have a very different idea. They have examinations that students can study for and teachers can teach to. Exam questions and examples of good responses are released each year after the exam. This is not only much fairer than what we do. It is also the best way to make sure

BEST COPY AVAILABLE

that the schools will develop curricula and teaching techniques that will actually prepare their students to succeed on the examinations.

Last fall, the National Center on Education and the Economy joined forces with the Learning Research and Development Center at the University of Pittsburgh in a development effort that could lead to a national examination system for the United States. Dr. Lauren Resnick, who is also a member of this panel, is co-director of the Pittsburgh Center, and, with Michael Cohen of our organization, co-directs the New Standards Project. The John D. and Catherine T. MacArthur Foundation and the Pew Charitable Trusts have provided $2.5 million for the first 18 months of this effort.

We are now assembling a group of some twenty states and school districts that will join us in the development of an examination system. Their leaders will set policy for the exams, deciding on the subjects that students will be examined in and setting the standards to be used. It is very important that these standards be set at internationally competitive levels. We believe that a group of volunteer states and school districts committed to setting such a standard are more likely to succeed than all 50 States and territories acting together.

We do not, however, see ourselves as simply setting out to construct a European-style examination system for the United States. There are some features of those systems that we would not want in the United States and many features on which we believe we can improve.

Dr. Resnick's testimony contains the outlines of our design. With your permission, I would like to include in this hearing record a copy of an overview of our project. My purpose here is to highlight certain principles that we believe are particularly important for any examination system that might be adopted in this country. They are:

—*Use the examination system to set a high standard that all students are expected to meet, not to sort them into various levels of competence.* Some may take longer than others to meet the standard, but it is vital to create an environment in which all kids are expected to learn.

—*Create examinations that students can study for and teachers can teach to.* The system should also be "open"—most items and examples of good responses should be released each year after the exam. This is the fairest way to run the system and the best way to make sure that the schools will make the major improvements in curriculum and teaching techniques that are needed to achieve the national education goals.

—*Include assessments of student performance on big, long-term tasks that cannot be included in a timed performance exam.* This is the only way to measure students' capacity to use what they know to do "real world" tasks.

—*The burden of reaching the standard should not fall exclusively on the student.* Embed the examination system in a program to restructure the schools. This involves pushing decisions about how to help students meet the standards down to school staff, and rewarding that staff when their students succeed and providing consequences for them when they do not. It also means using the standards to guide the professional development of teachers. And the teachers in our schools must have the access to information and staff development resources they need so they can develop local curricula that will enable their students to reach the standard set by the examinations.

—*Design the system to produce both excellence and equity,* not just excellence. It is easy to design a system that simply sorts kids out and establishes which ones make it to the top. The challenge is to create a system that gets almost everyone to a high standard.

This last is critically important. The whole American education system is predicated on the expectation that only a few will succeed at high levels and education performance overall will be distributed over a curve. The Nation's testing systems reflect those expectations; their very design requires the assumption that large numbers of students will perform badly. This is precisely the opposite of the premises of this project. Our object is to make it clear what students need to know and be able to do in order to succeed in school and life and then organize the whole system so that almost all students can achieve that standard. It may take some students longer than others, but that is all right. This is the only kind of system that makes sense if we really believe that all kids can learn. For that reason, in our view, it is the only kind of system that offers any real hope to minority and low income students.

As matters stand now, most minority and low income students are implicitly held to a lower standard than others from the middle of elementary school on, because

less is expected of them. An examination that simply sorts students by measured achievement will simply reinforce the inequities that already exist.

In the system we have in mind, the students, their parents and their advocates will be armed for the first time with clear information about what their children have to do to succeed, clear information about how well they are doing against that standard as they progress through school and a clear standard by which to judge the adequacy of the school curriculum. Students, for the first time, would not bear the whole burden of meeting that standard, because the professionals would be held accountable for student progress. The measures themselves would not be limited to timed performance tests, but would include opportunities for students of many different cultural backgrounds to select their own projects and tasks to demonstrate mastery, making it possible for them to choose performances that play to their strengths. Students who need more time to reach the standard would, as we have said, have the time they need.

Minority and low income students have been very badly served by a system in which very few people expect them to do well, and so reward them for what is in fact mediocre performance. The challenge is to create a system that sets a high standard and is structured to enable almost everyone to reach it.

The system we have in mind meets these objectives. We believe that it is both necessary and wise to view NAEP as a complementary monitoring system. And we also believe that a good national examination system should embrace a number of different examinations, not just the one we are working on.

The design we are using should produce an examination system for the core subjects—reading, writing, speaking, listening and mathematics—within 3 years. Within 3 years after that we believe examinations for all the subjects at three different age levels could be fully field tested and available for use.

Our project is a private undertaking, involving foundation funding and the voluntary participation of a number of states and school districts, as well as many teachers, researchers and experts in a wide variety of fields. The question is, what role should the Federal Government play in establishing a national examination system in the United States?

You may be able to help get the process for arriving at a national consensus on standards going. You certainly can make sure that there are sufficient funds for research and development on student performance assessment so that examinations can be prepared that measure what the country wants to have measured. And you can also provide some resources that will help make it possible to put a system in place so that the examination system will not simply measure failure, but instead help all students to meet high standards. I will expand on each of these points below.

The steps toward a national examination system begin with development of a national consensus on a framework for defining what students should know and be able to do. The National Education Goals Panel seems to be headed toward engaging in such a process. Whether the panel will need some form of Congressional authorization to do so will depend on the role the administration wishes to take, how the panel organizes itself for the task and the scale of resources needed But, if such Congressional action is needed, I hope it will be forthcoming.

In the meantime, we and others will be working in parallel fashion on such a process. In my view, the Nation would be well served by having several such efforts in motion at once. One reason is that the National Education Goals Panel will necessarily have to engage all the states in the consensus-building process. There may be a lot of pressure on the panel to come up with a standard at or below the average of the current state standards. It is, in our view, essential that the United States create an examination system that embraces an international standard. It is more likely to do that if one or more of the examination projects is free to show what an examination benchmarked to international standards might look like.

While the country is working toward a consensus on standards, it will pay, I believe, to have several efforts going forward at once to develop actual examinations. Both we and the chairman of the National Education Goals Panel hold the view that the country will be best served by having not a single examination but a system that makes it possible for individual states, districts and even schools to choose from among many examinations calibrated to a single standard.

If the Congress shares that view, then the wisest course may be to authorize the Department of Education to make funds available for competitive grants to organizations interested in developing methods for calibrating examinations to one another and to develop new examinations. The quality of the examination system eventually embraced by the United States will depend greatly on the extent and quality of the education research and development done on examinations over the

next few years. I would strongly urge the Congress to consider the principles outlined above when authorizing such research support.

An examination system seen as a silver bullet—a strategy which by itself will solve the problem of poor educational performance—is bound to fail. If the only change we make is creating the examination system, then we will simply have a better mechanism for measuring and penalizing failure; we will not have improved the prospects for millions of our students, especially poor and minority children. Legislation designed to put the other needed system elements in place is needed now, well before the examinations are ready for widespread use.

I therefore urge you to consider authorizing the Office of Research and Improvement in the Department of Education to support a wide variety of research, development, demonstration, evaluation and dissemination efforts designed to put in place the elements needed to make sure that the examination system produces the effects we all hope for. This program should be designed to lead to major advances in site-based management, new incentives for professionals, local curriculum development, new approaches to staff development and teacher training, and much more—all keyed to the new standards developed by the consensus process.

The one thing I would not do now that some might urge upon you is to create and charter a national examining organization. That, in my judgment, would be premature Doing so now might actually make it harder to achieve the goal of national examinations. There is great resistance in this country to anything that looks like an examination imposed by the Federal Government on a country historically suspicious of any effort by the Federal Government to impose national curriculum on the states and localities. No matter how serious an effort was made to create a National Examination Board that was independent of the Federal Government, the very fact that it was initiated by Congress would put it under a cloud.

My guess is that 2 or 3 years from now, when it may be necessary to have a single entity that administers a national examination system, a consensus will have emerged as to how to establish such a body, and Congress will be able to do so with the support of almost everyone whose support is necessary.

Decades of inaction on the sorry state of American education have been replaced by a sense of national urgency long overdue. I am convinced that, if the country takes the course just recommended, we can have both a National Examination Board and an initial set of exams in place within 3 years that will serve as the base for a system that will serve the country well for a long time to come.

Senator PELL. Ms. Daniels.

Ms. DANIELS. Mr. Chairman, members of the subcommittee, I am Maureen Daniels, a 6th grade teacher at Lynbrook Elementary School in Fairfax County, VA, and president of the Fairfax Education Association.

I appreciate this opportunity to speak with you on behalf of the National Education Association about the issue of student testing.

I would like to share with members of this subcommittee the experiences of a practicing teacher. It is no mistake that the chief proponents of the national standardized tests are not themselves teachers. Politicians, assessment experts, education researchers, and others speak as if more testing were the answer. Well, if testing is the answer, what is the question?

Practicing teachers are well aware that students are already bombarded by a plethora of standardized tests required by authorities which are in addition to other assessments needed to monitor students' daily progress. The 6th graders I teach endure 15 hours already of standardized tests each year. These layers of tests and assessments reflect the complicated and, at times, confusing layers of authority that govern our schools.

Those of us in the classroom must daily sort out competing and often conflicting demands of State and Federal regulations, district and school building policies, the appeals of individual parents, and our own sense, based on education and experience, of what is right and appropriate.

Our commitment to working within that environment must acknowledge that while complicated, the American public education system has two unique strengths: Diversity, and its ability to give students a second chance.

During the last 10 years in Fairfax County, our minority enrollment tripled from 9 to 27 percent. By the year 2000 it will reach 42 percent.

In 1992 Fairfax County public schools will have approximately 6,300 students from 150 countries speaking over 100 languages. For many of these students, this has been their only chance to receive an education. But national student testing would undermine both of those strengths, diversity and giving students a chance.

Teachers are concerned that a single national standardized test would lead inevitably to a national curriculum. And yet, our funding and policy mechanisms are all designed to provide State and local authorities a high degree of autonomy.

Despite rhetoric that testing would only establish goals and allow individual schools and teachers to determine the best way to achieve those goals, we must be realistic. If the national test is imposed, a test with significant consequences in funding, personnel decisions, and curriculum, virtually every hour of every school day will be spent preparing students to perform well on that test.

Many proponents of the national test point to the efficacy of national testing in Japan and our European economic competitors. I urge those testing advocates to take a careful look at the impact that that national test has on the students themselves.

In America, individuals have a wide array of academic choices they are entitled to make. They can choose among a vocational, basic education, or a college preparatory track. They can change their minds and attend community college or a university. They can change their minds again and return to college or an adult education vocational program. And then if they wish, they can change their minds again.

In Japan and in Europe, tests determine for individuals what courses they take at the secondary level and what their postsecondary educational options are. In short, a national test determines their future.

Make no mistake, opposition to national testing is not equivalent with opposition to holding students, teachers, and schools to high standards. Teachers embrace high standards and support effective student assessment.

But teachers and parents understand that assessment must be considered in a broader context. Some skills cannot be measured by a pencil-and-paper test. Tests often reveal more about test-taking skills than they do about knowledge. And it is extremely difficult to eliminate cultural and gender biases from many standardized tests.

Let me make one last point. It is reasonable to conclude that a national test would reveal that some students do better than others, some teachers do better than others, some schools do better than others.

But I believe any valid national test would reveal that schools perform best when they have sufficient resources to attract and retain quality staffs, have a broad academic program, and adequate

equipment and materials. Any valid national test would reveal that teachers do best when they have solid preparation programs, quality in-service education, and support from parents and the school itself.

Any valid national test would reveal that students do best when they come from families who value education, who have the resources to provide adequate nutrition, health care, and other basic human needs for their children. That is where our focus should be: On ensuring initial success for our children when they first come through the schoolhouse door. That is where we can really make a difference.

I sincerely believe your time and mine would be better spent in efforts to achieve that goal than in debating the value of a national test. I urge you to join me and the other NEA members in doing just that.

Senator PELL. Thank you, Ms. Daniels.

[The prepared statement of Ms. Daniels follows:]

PREPARED STATEMENT OF MS. DANIELS

I am Maureen Daniels, a sixth grade teacher at Lynbrook Elementary School in Fairfax County, VA, and president of the Fairfax Education Association. I am pleased to speak to you today on behalf of the National Education Association which represents more than two million education employees in the Nation's elementary, secondary, vocational, and postsecondary education institutions.

NEA has long opposed a single, standardized national test for students. Already, students from kindergarten on are subject to a broad array of standardized—tests required by local or state policy—and of teacher-developed and administered tests. Adding another test to the mix would do nothing to advance any significant education goal, and, in fact, could undermine other important education objectives.

NEA believes that—in the context of efforts to achieve the national education goals set by the President and the Nation's governors—it is appropriate to examine the whole area of assessment and accountability, and we agree that students, teachers, and schools should be held to high standards. We simply do not believe that the imposition of a national test is consistent with these aims.

OVERVIEW

The National Education Association recognizes the need for ongoing comprehensive evaluation of student progress. We believe that student learning must be assessed with measures directly linked to the lessons teachers teach and the materials teachers use.

Student assessment should not be equated with testing, and certainly assessment should not be equated with standardized testing. Appropriately, student assessment should involve teacher-developed and administered tests, formal and informal observation, and it should include assessment techniques that measure a broad range of skills—many of which cannot be ascertained through a pencil and paper tests. Student projects, oral reports, and other assessment methods must be used to demonstrate mastery of subject matter. None of these techniques alone should be used as the single determinant of a student's future—academic or otherwise.

Repeated studies demonstrate that standardized tests have inherent cultural and gender biases. All evaluation methods should be rigorously tested to eliminate such biases.

Opposition to national student testing should not be equated with opposition to holding students, teachers, and schools to high standards. NEA supports high standards and believe that local, State, and Federal officials who share responsibility for educational policies must examine those standards on an ongoing basis to determine whether they are in accord with the current economic and political environment and with national and local needs.

Moreover, NEA recognizes that there is a place for standardized tests. But in order to assure that such tests are used appropriately, standardized tests must be developed, administered, and interpreted by trained personnel. Further, the results

of those tests must be returned to the teacher and other appropriate staff in a timely manner in order to modify the instructional program.

NEA supports truth-in-testing as an important step for bringing about test reform, and we urge adoption of procedures that allow test takers to request a post-test review of all standardized test questions, scores, and rationale for the correct answers.

National Student Testing

The current drive to establish a national student test threatens to undermine the two chief strengths of the American system of public education: Its diversity and its ability to give students a second chance.

Many proponents of a national test argue that a national test would merely set goals for a body of knowledge, and that local and state officials would have a high degree of discretion and autonomy to determine how best to achieve those goals. But the fact is that a single, national student test would lead inexorably to a national curriculum. If a national test is established—with significant consequences for funding, personnel decisions, and curriculum—then virtually every hour of every school day will be devoted to preparing students to do well on that test.

Not only would this tend to undermine the diversity of our public school system and the authority state and local officials now have, but it runs directly counter to the goal of better integrating higher order thinking skills—such as problem-solving, analysis, and synthesis—into the curriculum. Most nations which have a standardized national student test tend to use their educational systems to discourage diversity. While Americans investigate why it is that Japanese students, for example, tend to perform better on standardized tests, Japanese educators are doing research in this country to find ways to encourage greater diversity and creativity in their educational systems.

Moreover, nations that use a national test generally use those tests as a gatekeeper. A student's performance on tests at the secondary and postsecondary level can determine, once and for all, their educational and economic future. The United States places a high priority on individual choice and on giving individuals a second chance. In consultation with parents, teachers, and counselors, a secondary student in a public school in this country has a number of choices as to whether to pursue a vocational track, a basic educational track, or a college preparatory track. In Japan and most European nations, those decisions are determined by a student's performance on a test.

In the same way, Americans have a wide array of options in postsecondary education opportunities. In the United States, an individual may choose to attend a college or not, may go back to school after a short or long hiatus, may pursue postsecondary vocational education or an academic education. As a percentage of the total population, more Americans attend postsecondary education institutions than any other nation. We make extraordinary efforts to provide that opportunity to individuals because we recognize that people's interests and level of commitment may change over time. A national student test would tend to diminish the ability of our society to provide those opportunities.

Improving Student Assessment

NEA strongly supports efforts to improve student assessment as a means of providing useful information about the effectiveness of educational practices and policies. Meaningful assessment methods should allow educators to diagnose student learning needs and select appropriate instructional strategies, as well as provide feedback to students and parents. They should enable school staffs to determine the effectiveness of curriculum and textbooks. And they should inform the community about how their schools are doing and what the schools' needs are.

If one can accept that these are the primary purposes of student assessment, it should be clear that traditional standardized tests have limited use for the changing needs of education. Rather than establishing a national test, Congress should support inservice education in evaluation techniques, research in effective assessment methods, and other programs that enable school faculties to improve the local assessment and, as importantly, to act on the conclusions of that assessment.

Conclusion

NEA believes that a national student test would only confirm what we already know—that some students do better than others, some teachers do better than others, and some schools do better than others. NEA believes that a national test

would reveal that schools perform best when they have sufficient resources to attract and retain quality steffs, have a broad academic program, and adequate equipment and materials. It would reveal that teachers do best when they have solid preparation programs, quality inservice education, and support from parents, the school itself, and ancillary agencies in the community. It would reveal that students do best when they come from families who support education and who have the resources to provide adequate nutrition, health care, and other human needs.

Achieving that difficult agenda—assuring equity of educational opportunity through adequate resources and support from the community and the society at large—is the only way to achieve the national goals in education established by the President and the Nation's governors. NEA believes that we have sufficient information to know what it will take to achieve those goals. All that remains is for us, as a Nation, to provide the resources, programs, and policies to achieve them.

Thank you.

Senator PELL. Mr. Faldet.

Mr. FALDET. Thank you, Mr. Chairman and members of the committee. My name is Burt Faldet. I welcome the opportunity to appear before you representing the Association of American Publishers. I am president of Test Consultants, Ltd., which provides evaluation, design, and implementation strategies to education and business.

Senator PELL. Could you pull the mike a little closer, please?

Mr. FALDET. I certainly can. Thank you. How is that? Better.

I have also been a high school teacher, a school psychologist, and I am now teaching graduate courses to teachers about how to test and develop good tests.

Quite frankly, the task of developing a national testing system is secondary in importance and, I would guess, difficulty to the tasks of developing a consensus on what every child should know in any particular subject area and in any particular grade level and assuring that the national test will be administered and the results interpreted and used in a standard, bias-free way.

The first, and what I hope will be the most important message I leave you with today, is that developers and publishers of standardized tests should be seen as part of the solution for improving the quality of educational instruction, not as part of the problem. Publishers are an integral part of the educational system, providing an essential delivery system as well as taking the initiative for and bearing the risk of developing new and innovative materials.

The second message is that test diversity and competition should be encouraged to assure improved education and improved assessment instruments. Different objectives are served by different kinds of tests, and I can't help but reinforce the statement of Ms. Daniels that the important testing, most important testing, has to be done in kindergarten, preschool, 1st grade, because that's when you find out where the student is and are able to make those initial important instructional strategies work.

Different objectives are served by different kinds of tests. No one test can accomplish all the diverse objectives of our education system. It is a serious mistake to try to make tests do what they are not designed to accomplish, and I think that is where a lot of concern about multiple-choice or true-false or essay tests have entered this arena.

It would also be a mistake to have only one test for a national testing system.

Finally, I want to assure the committee that test publishers, working with the education community, are expanding and improving their testing products to meet continually emerging educational demands. We are ready and willing and able to develop and administer whatever assessment instruments are desired, including assessments to measure progress toward attainment of national education goals or a school-leaving examination.

We urge that the development of any national testing programs be open and available to commercial test publishers and that both tests and instructional materials publishers be included at critical points in the formulation, development, and delivery of appropriate multiple-assessment instruments.

I think it is important to emphasize that not just basic skills but higher-order skills can indeed be measured, even in a multiple-choice format, in a standardized test. While we recognize that there are more direct ways of measuring this, these instruments have been available from publishers for some time. I would like to emphasize, too, that when you compare the relative costs of forms of measurements and the kinds of problems in the development of bias-free scoring, you are talking about an increase in cost of, conservatively, 10 times over what we are now doing.

What recommendations do we have for Congress? Well, the first is that certainly you continue holding hearings such as this on education issues, particularly testing, as a prelude to any possible future action. We feel that you should assure diversity of testing. No single test, no single curriculum, no single textbook can or should meet our Nation's diverse educational needs.

Third, developers and publishers have a role in making whatever testing program is adopted work, whether at the school level or for national education goals programs or for a school-exit program. Publishers and developers provide an economical and efficient delivery system for assessments. Publishers have traditionally served as an important bridge between sound theory and sound practice. Indeed, they have been the vehicle for getting local school acceptance for new concepts and the resulting products and for enhancing and modifying those products, as needed.

One of the crucial concerns is the proper interpretation of test results. A recommendation we would make is that Congress consider funding for targeted in-service training to teachers and administrators in interpreting test results to enable them to use tests better to improve instruction and to convey information to students, parents, and the public.

Finally, we do not believe that the Federal Government should get into State and local testing business. Any national testing program should not be a Federal program. The Nation's publishers commend the efforts to develop national education goals and stand ready to develop and administer the assessment instruments required to measure progress toward the attainment of such goals.

Thank you very much.

Senator PELL. Thank you very much indeed, Mr. Faldet.

[The prepared statement of Mr. Faldet (with an attachment) follows:]

PREPARED STATEMENT OF MR. FALDET

Mr. Chairman and members of the committee, by name is Burt Faldet. I appreciate this opportunity to appear before you today on behalf of the Association of American Publishers. The Association of American Publishers ("AAP") is the principal trade organization representing more than 235 member firms that publish hardcover and paperback books; professional, technical, and scientific journals; computer software; and classroom and educational materials, including tests and evaluation and scoring materials.

I am president of Test Consultants, Ltd, which provides evaluation, design, and implementation strategies to education and business. Our clients have included commercial test publishers, the American Institutes for Research, IBM, as well as individual school districts. From 1965 to 1987, I was with Science Research Associates, a commercial test publisher, where I was involved in a variety of positions in the development, publication, and use of standardized tests for schools and industry. I also have taught undergraduate courses in Measurement and Evaluation and secondary school science, and served as a School Psychologist and Director of Pupil Personnel Services.

There are several points that I would like to discuss today, from the perspective of the publishers of such tests, about the development and use of standardized tests in elementary and secondary schools—whether for the purpose of evaluating progress toward national education goals or to an individual pupil's development. My statement does not address higher education, employment, or military testing.

The first, and what I hope will be the most important message I leave with you today, is that developers and publishers of standardized tests should be seen as part of the solution for improving the quality of educational instruction, not as part of the problem.

The second message is that test diversity and competition should be encouraged to assure improved education and improved assessment instruments. Different objectives are served by different kinds of tests—no one test can accomplish all of the diverse objectives of our diverse educational system. It is a serious mistake to try to make tests do what they are not designed to accomplish, or to use tests as the sole means for assessment in most situations.

Finally, I want to assure the committee that test publishers—working with the educational community—are expanding and improving their testing products to meet continually emerging educational demands. Test publishers are ready and able to develop and administer whatever assessment instruments are desired, including assessments to measure progress toward attainment of national education goals.

WHY TEST?

Measurement can be relatively exact—but a number has no meaning until someone makes a judgment about it. That is the difference between measurement and evaluation. There are many ways to determine health; a number on a thermometer is one indicator, but it takes someone to exercise judgment as to the significance of the temperature shown, and to take the appropriate action as indicated by the reading on the thermometer. It would be imprudent, however, to rely entirely on temperature to make a diagnosis of the patient.

Why educational testing? Testing is of value to the student. It serves to provide some information that can be used by educators and parents to identify and respond to the instructional needs of individual pupils and to improve instruction of individual pupils. Testing is a means to assess progress toward specific educational objectives, as evidenced by what pupils can do in terms of skills exhibited.

Testing also serves broader, institutional goals. It assists in assessment of long-range effects of changes in the educational program, enabling comparison of (1) performance over time and to changes in the instructional program or to changes in population characteristics and (2) performance across different subject areas, such as mathematics and reading, to determine strengths and weaknesses, needs for program modification, or changes of emphasis. Testing is one means to evaluate performance for accountability purposes.

The methods of evaluating whether children are learning what is being taught have changed over the years, just as many techniques and objectives of teaching have changed. For example, standardized achievement tests and numerous other types of tests have supplemented teacher-made tests administered on a class-by-class basis; performance and portfolio tests are serving to complement and supplement multiple-choice formats.

LIMITS TO TESTING

It must be emphasized, however, that there are limits to testing. When testing is used in "high-stakes" situations and results are used as a simple "pass/fail" barrier to students, or to reward or punish teachers and administrators, when the pressure becomes so intense that there is "teaching the test" rather than teaching the skills and concepts that are being evaluated, when test scores become the sole criteria for evaluating student performance or potential or the effectiveness of instruction, then testing has gotten out of hand and is being misused and abused.

Tests are a necessary but not sufficient means to assess achievement and growth in skills and abilities. What may be tested is not, and cannot be, inclusive of all of the desired outcomes of instruction.

Tests may be used as a partial basis for evaluation. Tests are concerned only with certain basic skills and abilities and are not intended to measure total achievement in any given subject or grade; they are not inclusive of all the desired outcomes of education. Standardized tests are concerned with only those areas of instruction that are amenable to objective measurement.

It should also be recognized that local performance is conditioned by many influences. The instructional effectiveness of the teaching staff is only one of these factors. Among other factors are the pupils' school and home environment, their past educational history, and the quality and adequacy of the instructional materials with which the staff has to work.

As stated in the *Manual for School Administrators* for one standardized test:

> At all times, the tests must be considered a means to an end and not ends in themselves. These tests have their principal value in drawing attention of the teaching staff and the pupil to those specific aspects of the pupil's development most in need of individual attention; in facilitating remedial and individualized instruction; in identifying those aspects of the whole program of instruction most in need of increased emphasis and attention; and in providing the basis for more adequate educational guidance of the individual pupil. If properly used, the results should motivate both teachers and pupils to increased, better-directed efforts in both teaching and learning.
>
> When intelligently used in combination with other important types of information, the results obtained from these tests should prove very valuable in the appraisal of the total program of instruction. Unless they are used in conjunction with other information, however, they may do serious injustice to many teachers and to many well conceived instructional programs.

KINDS OF TESTS

Different tests have been developed to meet a variety of purposes. Some tests are subjective, both as to the matter tested and the interpretation of the results. A standardized test is an objective test that uses the same standards to measure and score student performance across the country; everyone takes the same test according to the same rules.

A normed-reference test (NAT) is a standardized test used to compare students' performance in terms of a carefully selected, nationally representative group, or norm, on the same test; performance is based on total test or subtest scores. (In contrast, for some tests, such as the SAT's and ACT, the norm is based on the others taking the test, rather than to a standardized national norm.)

A criterion-referenced test (CRT) differs from a normed-reference test primarily in how test scores are interpreted and used. A criterion-referenced test is used to evaluate and report performance in terms of specific instructional objectives or skills, stated in measurable terms.

These labels are not mutually exclusive. Many criterion-referenced tests are normed, and many norm-referenced tests may be subject to criterion-referenced, content-based interpretations.

Teacher-made tests generally are intended to provide information about individual student's performance on specific, classroom-oriented, curricula or specific needs for information about students. These tests are frequently supplemented by textbook tests, which are developed by textbook publishers and may appear in textbooks or be provided to teachers as supplementary instructional materials. Both of these tests are associated frequently with grades on report cards and help measure a student's progress in class, as well as facilitate individualized instruction.

Tests can also be in a variety of formats. Multiple-choice tests offer the advantages of objectivity and uniformity of scoring, ease of administration and scoring,

and low cost. There are disadvantages to such tests, particularly if they are utilized as the exclusive method of assessment. "Performance-based tests," "authentic assessments," or "alternative assessments" generally are open-ended tests that are not multiple-choice. They include essays, writing samples and portfolios of work, practicums, or oral or visual demonstrations. They generally are more expensive, labor-intensive, and require more training and preparation to administer and evaluate—factors which also can make them affirmative educational tools. The same concerns for fairness, validity and reliability, standardization if used for comparisons, and abuse if used in high-stakes situations that are raised with multiple-choice tests are applicable to performance tests.

Performance testing and standardized testing are not mutually exclusive. It is important to point out that for several years writing and listening assessments—performance tests—have been offered by test developers as part of their standardized test batteries. Publishers are now offering performance and portfolio tests to supplement their current test batteries.

What are the particular advantages of a norm-referenced, standardized test? It ensures reliability and validity in data collection, analysis, and interpretation. It enables evaluation of student achievement in various grades and subjects for the purpose of aggregating and reporting achievement gains in terms of a common reporting scale (e.g., normal curve equivalent or grade equivalent), with nationally representative norms. It provides an objective, rather than a subjective, assessment. The efficiencies are greater and the costs of administration are far less than for a performance assessment (by perhaps, a factor of 10).

Norm-referenced, standardized tests also enable identification of problems in specific skill or subject area deficiencies for teacher attention and remediation. This may be particularly important in the early grades.

Norm-referenced, standardized tests use the same or parallel test items for all students, which makes scores for all students comparable; use of one level per grade facilitates criterion-referenced interpretation of results for classes, buildings, and systems. Individual scores can be related to comparable national norms. One skill can be compared to another on a pupil, class, building, or system basis.

A classroom may have such a wide range of skills that no simple test can be equally suited to the entire range of achievement; NRT's for different levels of achievement can be administered so that each pupil takes the level that corresponds most closely to the individual instructional objectives and levels of skill development.

ROLE OF THE TEST DEVELOPER AND PUBLISHER

Test developers adhere to strict standards, as developed by the American Psychological Association, the American Educational Research Association, and the National Council for Measurement in Education in the *Code of Fair Testing Practices in Education*, a copy of which is submitted for inclusion in the record. Demonstration of reliability and validity also must be provided to test users, showing that the test meets its intended purpose and its appropriateness for groups of different racial, ethnic, or linguistic backgrounds who are likely to be tested. Several books give in-depth, candid reviews of available tests, include the *Mental Measurement Yearbook*, published by the Buros Institute of Mental Measurements, while guides and evaluations are published by the ERIC Clearinghouse on Tests, Measurement, and Evaluation and by other organizations.

Standardized tests generally are professionally developed tests distributed by commercial test publishers; development may be by the publisher, educators or other non-profit organizations (under royalty or other forms of compensation), or by governmental entities alone or in cooperation with publishers (such as under the National Science Foundation's "Publisher Initiative").

The role of the commercial test publisher in test development is very extensive. Based on information from a variety of sources, including the educational community, the test publisher determines if there is a need for a test and whether it will be financially viable. If the answers are in the affirmative, a decision is made as to the type of test to be developed, i.e., a norm-referenced or criterion-referenced test, or a combination of the two. In addition to the type of test, the format (e.g., multiple-choice, true-false, performance assessment) also must be determined. Publishers also respond to test requirements of state and local education agencies. Test publishers will also respond to the demands and requirements for tests of national education goals.

Extensive research is required for "building" a new test or revising an existing test. Test items are written by educators and professional test item writers. They

are selected after extensive research on educational objectives; curriculum; goals, objectives, and standards; textbook and instructional material content; and what is to be measured and how. In general, tests follow and reflect curriculum. Error-free items must be developed that will withstand the scrutiny of hundreds of thousands of teachers and students over a long-period of time. Vocabulary and readability levels must be appropriate for the students to be tested. Items must also be free from ethnic, gender, or cultural bias.

At least one tryout to obtain data for standard item analysis and summary test statistics is needed. This data is used to select items with desirable characteristics. Typically, an experimental edition will contain at least twice the number of items required for the final test, to enable the publisher to reject undesirable items and still retain a sufficient number of items for a final test of suitable length. Items for a norm-referenced test will be rejected if too many examinees select the Answer. In a criterion-referenced test students are classified in terms of mastery/non-mastery, so items will be selected that will have a large number of correctly-selected answers.

Experimental test items are reviewed by educators and curriculum specialists and are then field tested with large numbers of student to check their responses. The comments of the reviewers and the data generated by the field test are used to select the items for the final edition of the test.

In the case of norm-referenced tests, the final, or standardized, version of the test is administered to carefully selected groups of students whose characteristics are similar to those of students throughout the Nation. The information obtained is then aggregated into norms so that individuals tested in the future may be compared to the original national sample. This is the process of standardization, and the normative information obtained from the process is crucial to educators, parents, and students. Without it, there would be no way of knowing how a single score on a specific test compared to the scores of other students in the Nation.

Publishers develop guidance materials to assure that the final test is administered in accordance with the standardization, and to provide instruction on how the test is to be interpreted. Information is also developed and provided on the technical characteristics of the test to support its reliability and validity. This is done regardless of the format.

Scores can be reported and evaluated in a multitude of ways, for different uses. Rather than trying to describe scoring and interpretation in my testimony, I am submitting for the record an excerpt from *Understanding Achievement Tests: A Guide for School Administrators*, published by the ERIC Clearinghouse on Tests, Measurement, and Evaluation, on "What Types of Test Scores Are There."

Much controversy has been generated recently over norm-referenced testing. To address these concerns, I am attaching to this statement several articles from commercial test publishers that were included in the Summer 1988 *Educational Measurement: Issues and Practice* that provide an extensive review of these issues.

WHAT SKILLS ARE TESTED?

Higher order skills, not just basic skills, can be measured, even in a multiple-choice format, in a standardized test (remembering that it was only a very few years ago that publishers had to respond to demands for assessment instruments for the "back to basics" movement). We recognize, too, that there are more direct ways of measuring higher order skills and such instruments have been available from test publishers for some time.

As previously stated, the multiple-choice format used in assessment instruments has some attractive features. It is an efficient and effective way of measuring many educational objectives. While we recognize that it has limitations as well, it is important to recognize that most measures, including criterion-referenced and performance tests, are samples of behavior from which inferences can be drawn. For example, a multiple-choice mathematics test, which includes five exercises in addition of two-digit numbers with carrying, is a sample of all the possible two-digit numbers that we want a student to be able to add. For efficiency, we chose five exercises, and based on the student's performance on those, we infer what the student could do if presented with many more. Similarly, we may present a situation with several complex problem-solving exercises in a multiple-choice format. Based on performance, we can make some inferences about the student's performance in some of the higher order skills in the mathematics area.

Similarly, we can infer some important aspects of performance in writing from items commonly presented in multiple-choice language arts tests.

Neither the problem-solving nor language arts tests are substitutes for direct observation of student performance over time and in different situations in solving problems and in producing written material.

Reiterating a constant theme of this statement, that tests need not be mutually exclusive, I again want to point out that publishers of standardized tests currently also offer a variety of performance tests (including listening skills and writing) as well as portfolio programs, in addition to multiple choice tests.

Whether multiple-choice or performance tests, the keywords for the future, as they are today1 are validity and reliability. Publishers cannot and should not market a test unless it has been demonstrated to be valid and reliable. This requires time and money, extensive research and development, testing and reworking to assure that the test works.

RECOMMENDATIONS AND SUGGESTIONS FOR FEDERAL POLICY

On behalf of the publishers of standardized tests, I welcome this opportunity to meet with the committee and discuss standardized tests and our role in the educational process. As I said at the beginning of my statement, publishers want to be part of the solution, not part of the problem. Publishers are not simply printers, bookbinders, and marketers. They are an integral part of the educational system, providing an essential delivery system as well as taking the initiative for and bearing the risk of developing new and innovative materials. Just as Congress would not think of addressing the future of the automobile without consulting with automobile manufacturers, publishers should continue to be consulted and included in your continued deliberations over the quality of education and the development and assessment of national education goals.

What recommendations do we have for Congress? The first is that you continue to hold hearings such as this on education issues, particularly testing, as a prelude to any possible future action.

Second, Congress should continue to assure diversity of testing. No single test, no single curriculum, no single textbook, can or should meet our Nation's diverse educational needs. Competition among test developers, including a vigorous private sector, should be encouraged.

Third, publishers have a role in making whatever testing program that may be adopted by a school or for a national education goals program work. They provide an economical and efficient delivery system for assessments. Publishers have traditionally served as an important bridge between sound theory and sound practice. Indeed, they have been the vehicle for getting local school acceptance of new concepts and the resulting products, and for enhancing and modifying those products as needed. They have been the primary link between those who create and those who must implement. We do not see a change in this role, nor do we believe that a change is desirable. For this reason, it is important to involve the publishers early in the conceptualization of products resulting from sound research.

One of the crucial concerns is the proper interpretation of test results. Our fourth recommendation is that Congress consider funding for targeted, in-service training to teachers and administrators in interpreting test results to enable them to use tests better to improve instruction, and to convey information to students, parents, and the public.

Fifth, state and local education agencies might be required to develop a comprehensive assessment plan, which would identify instructional and accountability goals and objectives and the assessment instruments that would be used to achieve them and measure progress. The plan could include specific programs for in-service training, public information, and for assuring that tests are selected, used, and interpreted appropriately.

Finally, we do not believe that the Federal Government should get into the state and local testing business. We commend the efforts to develop national education goals, and stand ready to develop and administer the assessment instruments required to measure progress toward attainment of such goals.

I would be remiss if I did not point out that while publishers are trying to respond to the need to develop challenging and innovative tests (parallel efforts are being undertaken by publishers of textbooks and other instructional materials), Federal tax policy is frustrating its achievement.

The Department of the Treasury is insisting that publishers of tests and instructional materials capitalize research and development and other pre-publication costs, a position that falls with special weight on preparation of new tests and instructional materials, with their high development costs, high risks, and long lead times. This approach is shortsighted as a matter of educational policy because it dis-

courages the development of the innovative quality tests and textbooks our schools need. It is also discriminatory and unjustified tax policy because it requires capitalization of product development and research costs that, for any other industry, could be deducted in the year incurred. We have requested the tax-writing committees (and the administration) to provide appropriate relief, but the outcome remains very uncertain. This committee's assistance in assuring that tax policy does not frustrate education policy would be most welcome.

Thank you for your attention. I would be pleased to respond to any questions the committee may have.

EDUCATIONAL MEASUREMENT
ISSUES AND PRACTICE
Volume 7, Number 2
Summer 1988

Riverside Comments on the Friends for Education Report

Edward C. Drahozal
Riverside Publishing Company
and
David A. Frisbie
The University of Iowa

The authors, both affiliated with Riverside Publishing Company, discuss the factors they think explain the Lake Wobegon phenomenon and call for more appropriate use of normative comparisons and more complete reporting of test results.

The final report of Friends For Education (FFE), "Nationally Normed Elementary Achievement Testing in America's Public Schools," has appeared in several preliminary forms, and portions of its contents have been discussed in several national newspaper reports. The following comments are based solely on the content of the final report, and, though we take exception to some of the data analysis methods employed, some of the interpretations offered, and some of the conclusions reached, these remarks focus primarily on questions raised by the report. These are issues that complicate the use of aggregated test scores—with regard to both their formation and their interpretation—to synthesize state versus national comparisons in the face of different test batteries, different standardization samples and methods, different years of norming, and different score scale units for reporting.

The study of state and district performance reported by FFE appears to have been conducted as carefully as possible under the circumstances. The issues that the report raises—accuracy and comparability of norms, currency of norms, selectivity of pupils tested and reported in reports to the public, and temptation to teach specific content when educators are under accountability pressure—are not new. What comes as a genuine, unexpected, disappointing shock is the apparent universal appeal of the

Edward C. Drahozal is Senior Program Administrator at The Riverside Publishing Company, 8420 Bryn Mawr Avenue, Chicago, IL 60631. He specializes in practical aspects of interpretation, use, and reporting of elementary and secondary tests, primarily standardized tests.

David A. Frisbie is Associate Professor and Assistant Director of the Iowa Basic Skills Testing Program, University of Iowa, 316 Lindquist Center, Iowa City, IA 52242. He specializes in achievement testing.

simplistic objective of being above the national average and the extent to which schools are successful in managing somehow to appear above the national average when faced with pressures and even ultimatums from politicians, press, the courts, and even watchdog groups.

Despite the shortcomings that can be cited regarding the nature of the data that FFE analyzed and reported, there is ample evidence to warrant close examination of the group's fundamental question: Why are so many pupils or schools or states appearing to perform above the national average? The question seems as simple and straightforward as the one posed several years ago: Why are test scores declining from year to year? We believe that the question raised by FFE rivals the score-decline question in significance and, as was true of the score-decline inquiry, this search for resolution is likely to yield multiple, concomitant explanations. There is no single best answer. A closer examination of the issues by FFE, the publishers, and the state and district test coordinators might enhance our ability to use test data to further our primary goal—to improve the quality of instruction provided in our schools. With this purpose in mind, the remainder of this paper is devoted to identifying what we believe are the most crucial issues and to presenting a scheme that we would use to compare the performance of state or district groups with national pupil or school norm groups.

Some Major Issues

Accuracy of Norms

National norms for standardized achievement tests are based on a sample of pupils and schools (attendance centers) obtained through a complex, multistage sampling scheme. Each publisher strives to ensure that the national population of pupils and schools is properly represented in its norms sample. For example, in the standardization of the Iowa Tests of Basic Skills (ITBS) in 1984–85, districts were chosen on the basis of geographic region, enrollment size, and socioeconomic characteristics of the communities served. The standardization is a joint responsibility of the authors, publishers, and school personnel. Rigid conditions for district participation included the provision for sampling attendance centers of the district by the publisher rather than by the school administration. An adequate sampling plan is necessary but not sufficient to guarantee adequate norms. Only if the plan is realized, only if the sample obtained reflects the sample desired, will the norms represent national pupil or school achievement accurately.

To the extent that any publisher's norms misrepresent the national distribution of pupil and school achievement, comparisons with either of these norm groups will distort the estimated achievement level of the group in question. An underrepresentation of high-achieving schools or high-achieving pupils will cause the national norms to be "softer" than they ought to be. That is, an average-achieving pupil will appear to be above average when referenced to a group whose average is below their theoretical or "true" average.

The sampling plan, nature of the obtained sample, and weighting schemes used in the standardization of each achievement battery in question should be examined to determine the representativeness of the published norms. This should be done separately for pupil and school norms.

Recency of Norms

It is a well-documented fact that achievement in grades 3–8 has been rising steadily since the late 1970s. Though the year-to-year differences might be regarded as minor (.3 of a grade-equivalent month, on the average), the cumulative effect over 10 years is significant (approximately 3 months, on the average). Obviously, those who compare the 1987 performance of their pupils with that of other pupils who were tested in 1978 (national standardization) will be using "softer" norms and will have more pupils appearing to be above the national average than really are.

We have published information on changes in student performance for the past 30 years. Data for 1955 to 1984 are summarized on pages 148–153 of the new ITBS *Manual for School Administrators* (Hieronymus & Hoover, 1986). Differences in performance vary by test, grade, and score level. The 1977–85 composite score differences are eight to nine percentile ranks (PRs) at the median in most grades, but differences in language exceed 10 PRs in several grades at several score levels.

In periods of fluctuating achievement levels, the recency of the norms is a critical issue. When achievement levels are relatively stable over time, as they have tended to be at the grade K–2 levels, "old" norms do not interfere with score interpretations, assuming that we have curriculum stability as well.

Nature of Tested Population

If we have good reason to believe that pupils in a given state should have scores, on the average, below the national average, we must be certain to define the population for which we expect the prediction to hold. There are several related issues regarding this point with respect to the FFE data. If State X reports a mean normal curve equivalent (NCE) for 45,000 fourth graders, we should ask these questions: How many fourth graders were tested but not included in the computation of the mean, and what is the nature of the scores of those who were excluded from reporting? How many fourth graders are there in State X who were not tested and, consequently, who were not included in the reported scores? And what are the achievement levels like for these students who were not tested?

Based on the Department of Education's Center for Education Statistics fall 1985 enrollments projected to 1986, the percentage of students for whom scores are reported in the FFE report varies from a low of about 85% to more than 96% of total grade enrollments for most states for which full-grade testing was reportedly done. (For one state with public school enrollments of about 48,000 students per grade, averages and PRs are reported for approximately 37,500 students, which is about 80% of the total enrollment.) The discrepancy between the reported state scores and the expectations in the FFE report may be in part due to such differences between tested and total enrolled populations of students and

specifically to the nature of the portion of the student population not tested.

Adequacy of Expectations

Educators have developed some expectations about how pupils and groups of pupils might perform on achievement tests based on their study of the relationship of school achievement to other social, political, and economic variables. This is why we use such variables as enrollment size and socioeconomic status for stratified sampling in standardizations. FFE has used some of these relationships in attempting to develop expectations for state-level and school-district-level performance. Per-capita income, graduation rate, and college entrance score averages are among the "standard barometers of excellence" employed by FFE. Though we do not deny the value of these indicators as part of the prediction equation, we realize that it is not possible to predict achievement in this way with high accuracy. For example, the achievement test performance of Iowa pupils is among the very highest in the nation, yet these facts about educational conditions in Iowa seem inconsistent with that high level: Iowa ranks 27th among states in per-pupil expenditure, 39th in average teacher salary, and 44th in spending increase from FY86 to FY87.

In view of the less-than-perfect relationships between achievement and these other variables, the precision of whatever expectations about achievement we may formulate should be tempered. That is, what we are able to say with reasonable assurance about how many pupils or schools should score above a specific point (the mean in the national norms distribution) is not very useful. Consequently, we might instead settle for statements like these for State X: "About 40% of the fifth graders tested should score between the 25th and the 75th percentiles on national pupil norms," or "About 49–55% of the third graders tested should score above the national pupil median (50th percentile)." Of course, the ability to make such statements depends on a far greater understanding of the statistical relationship between those variables than most states probably have been able to determine.

Teaching the Test

Pupils and their teachers who participate in the standardization of an achievement battery have not had an opportunity to see or study the specific test questions used. Thus, having no practice on the specific test questions is one of the stringent preconditions of the standardization process. Subsequently, when these norms are used to interpret the scores of pupils who have been drilled with the exact test questions, the result is an overrepresentation of the amount of knowledge and skill possessed by such pupils. Likewise, when the scope of the curriculum is narrowed to encompass primarily the objectives measured by the exact test questions, the relative standings of the pupils who experienced the restrictive program of study will be overestimated.

No publisher condones this use of tests, and few teachers probably follow such abominable practices. Those who do are nearly always motivated by significant negative consequences associated with scores that might turn out to be below expectation (not always synonymous with national average). Unfortunately, for some educators, job retention and salary increases are tied directly to the test scores of their pupils. The authors of the ITBS have always decried the use of achievement scores for such purposes and instead have campaigned for the use of these scores to improve instruction directly.

If certain tests are to be used strictly for accountability purposes, their security must be ensured so that the scores that result will be valid for that purpose. The dollars required to assure states and districts that the test forms they will use are secure would be far greater than the value of the information derived from using the secure forms. Those dollars would likely have greater and more visible impacts on learning if devoted to direct instruction instead.

Score Analysis and Interpretation

With which norm group, pupils or schools (attendance centers), should averages from State X be compared to interpret the scores of pupils from that state? With which norm group, pupils or schools, should averages from District A be compared? There are only two choices, pupils and schools, because no publisher provides norms for school districts or for states. This is a fundamental issue currently facing the Council of Chief State School Officers as they contemplate options for providing for state-by-state achievement comparisons in the future. The choice to be made is not a matter of personal preference but a matter of the logical correspondence between the units to be compared. That is, averages of school buildings should not be referenced to a distribution of individual pupil scores, district averages should not be referenced to the distributions of either school building averages or pupil scores, and state averages should not be referenced to any of these three distributions. In view of the differences between these separate distributions, it is most logical to reference a score or average score to its own kind. When the most logical referencing is not possible, appropriate caution should be exercised.

The national pupil norm group includes pupils whose scores on a test are as high as perfect (PR = 99) to those whose scores are as low as zero or chance average (PR = 1). No school (building or attendance center) is likely to have an average score that is perfect or zero. In fact, on the ITBS and any other test with recent school norms it is reasonable to expect that no school will have a raw or scale score average higher than PR 88 or lower than PR 12 compared to the pupil distribution. Because many school districts are single-grade-within-single-building entities, the distribution of school district averages probably would encompass the same range as the distribution of school building averages. The school district distribution, however, is likely to be markedly more leptokurtic and less variable than the school average distribution. In terms of the pupil distribution, the distribution of district averages might range, *effectively*, between PR 75 and PR 25. Finally, most of the state averages on a test for a given grade might well have actual bounds that corre-

spond to PR 60 and PR 40 on the pupil distribution.

Because norms for district averages or for state averages are not available, districts and states often use the pupil and school norms that do exist. When a district average is referred to the pupil norms, it should be thought of as the score obtained by the average pupil in the district. We might find, for example, that the average pupil in District A scored higher than 63% of pupils nationally. Using the same rationale and the estimate given above, the average pupil in most states is not likely to exceed PR 60 or fall below PR 40. The value of such information is highly questionable.

A matter related to this general issue of analysis concerns the methods of computational precision used to aggregate and convert scores. As an example of the problem, a grade 4 school average GE composite score of 42.0 (obtained in the fall) on the ITBS has a PR of 46, and a score of 43.0 has a PR of 53. By interpolation and rounding, an average GE of 42.5 corresponds to a PR of 49.5 or 50. If GEs are rounded before converting to PRs, a 42.5 could be treated as a PR of 46 or 53, depending on the rounding convention adopted. Of course, this illustration underplays the magnitude of the distortion that could result with distributions of either school district or state averages.

Publisher-User Responsibilities

Though it is in the best interest of both publishers and test users to have tests and scores used properly, neither can ensure that the other will do its part willingly and unselfishly. Publishers must be counted on to standardize and analyze results in professionally acceptable manners. They must guard against potential misuse by informing educators of the intended uses of the tests they publish and warn against the possible misuses that might be anticipated. Publishers must do their utmost to provide test materials only to those who are at least minimally qualified to handle the tests and scores in a professional way. State directors, superintendents, teachers, school boards, and the public, generally, do not have the resources to monitor the effectiveness of publishers in attending to these obligations.

Publishers, on the other hand, cannot monitor the use of their instruments effectively to curtail misapplication, misuse, or misinterpretation. Often after the fact, a publisher can recognize inappropriate use—whether intentional or unintentional—and attempt to persuade the user to modify a proposal or report. Some school districts perform extensive audits to ensure that all students who were to be tested in each attendance center were actually tested. Some districts also audit results and retest suspect groups. But for the most part, publishers are not aware of and have no control over school districts' test administration conditions, the students included in summary data reported to the public, or methods used to synthesize data to make test results more palatable for less sophisticated consumers.

Most test authors and publishers go well out of their way to comply with the standards for educational and psychological tests adopted by the profession. Test score users—teachers, administrators, legislators, and other public groups—tend to know far less than they should about the nature of tests or the principles with which test makers intend for scores to be used. We should not denounce a test because a state committee uses the wrong norms or incorrect statistical analysis procedures in reporting. Likewise, we should not blame users for results based on shoddy standardization procedures or on inadequate or deceptive descriptions of such procedures.

Finally, publishers are obligated to clients to maintain the confidentiality of test data. It has been and should continue to be each client's decision to release test data and to determine the nature of any data to be released. Reporters, citizens, citizens' groups, and others who wish to obtain test data should respect this publisher-client relationship and seek release from the school district or state, depending on their level of interest and the dictates of state law.

TABLE 1

Percentages of State X Pupils Performing Within Selected National Pupil Percentile Intervals

National percentile rank	National percentage	Grade K	1	2	3	4	5	6	7	8	Average K–8
90–99	10	20	21	24	20	21	20	20	21	19	20.7
75–89	15	22	24	23	24	23	25	24	24	23	23.5
50–74	25	26	29	25	27	28	28	29	28	28	27.4
25–49	25	19	16	18	17	16	17	17	17	19	17.6
10–24	15	8	6	7	7	8	6	7	6	7	6.9
1–9	10	5	2	3	4	3	3	3	4	4	3.4
Percentage above national median		68	74	72	71	72	73	73	73	70	72
Percentage below national median		32	26	28	29	28	27	27	27	30	28

A Sample Reporting Method

We recommend an approach like the one described below for states that wish to describe the achievement levels of their pupils in relation to pupils in a nationally representative norm group. Exactly the same procedures could be used with school building (attendance center) data. Table 1 shows national PR ranges in the first column and the corresponding percentages in the second column. The body of the table shows, separately for each grade, the percentage of pupils in State X that obtained national PRs in each range. The last column shows the row averages of the percentage values. (Note that these are percentages and not percentile ranks and, consequently, it is acceptable

BEST COPY AVAILABLE

to average them.) The bottom two rows indicate, again by grade, the percentage of pupils above and below the national median. A histogram with one distribution superimposed on the other or a simple bar graph would provide a helpful visual display of the same information.

The main advantage of this method of reporting compared with reporting simply the percentage scoring above the national median is obvious. Between-grade differences and similarities can be examined, but most important, discrepancies from the national distribution can be accounted for in each of several segments of the distribution. If all we know is that 72% are above the national median, we do not know if the "extra" 22% are mostly located very near the median, mostly spread through the upper half, or mostly concentrated in the tail. Also, we do not know if the extra 22% are shifted from the lower tail, from throughout the lower half, or from just below the median.

Many districts use a reporting procedure similar to the reporting scheme described above. We recommend that such tabular data be supplemented with at least the following sorts of information: testing date, test form and level(s) used, type and date of the norm used, and percentage of eligible students tested.

Riverside Publishing Company and its representatives do not believe that the average pupil in every state has scores above the national median on the ITBS. We are confident in our standardization procedures and have subjected those procedures to public scrutiny in detail in the *Manual for School Administrators*. We have updated our norms at least every 7 years and, when achievement showed a pattern of increase in the early 1980s, new norms were obtained even though new test forms were not introduced. We are making plans to provide annual national norms updates for next year. Our manuals caution users about appropriate use of norm groups for varying purposes. Our hope is that the issues raised above will cause FFE and state and district test coordinators to reassess their analysis and reporting procedures to ensure that conclusions reached are based on a valid foundation rather than data of questionable origin and manipulation.

References

Hieronymus, A. N., & Hoover, H. D. (1986). *Manual for school administrators: Iowa Tests of Basic Skills*. Chicago: Riverside.

A Response to John J. Cannell

Joanne M. Lenke and John M. Keene
The Psychological Corporation

Two representatives of The Psychological Corporation present their reactions to the Cannell report and call for better explanations for the public of the meaning and limits of norm-referenced scores.

In recent years, public attention has focused on standardized achievement test results. These results, which are intended to describe the performance of individuals in relation to one another, are now often used to describe the performance of groups of students. In a report entitled "Nationally Normed Elementary Achievement Testing in America's Public Schools: How All 50 States Are Above the National Average," John Jacob Cannell attempts to cast doubt on the validity of the information being reported to describe the achievement of students as aggregated at the state and/or district level. The report states, "These tests allow all the states to claim to be above the national average! The tests . . . allow 90% of the school districts in the United States to claim to be above average. More than 70% of the students tested nationwide are told they are performing above the national average."

In response to Cannell, it is fair to say that many states and school districts report above-average performance in reading, mathematics, and/or language in the elementary grades. We do not believe that this is an attempt to misrepresent students' achievement in the nation's schools. Let us examine three very

Joanne M. Lenke is Vice President, Measurement, at The Psychological Corporation, 555 Academic Court, San Antonio, TX 78204-2498. She specializes in test development and norming, scaling, and equating tests.

John M. Keene is Director, Admissions/Credentialing, Customized Measurement, and Research, at The Psychological Corporation, 555 Academic Court, San Antonio, TX 78204-2498. He specializes in educational measurement.

important issues related to the interpretation of this information: (a) group performance relative to a national norm, (b) local performance relative to national performance, and (c) the stability of achievement test norms over time.

Interpreting Group Performance Relative to a National Norm

When a test is standardized, or normed, the test is typically administered to hundreds of thousands of students nationwide. This norming sample is drawn to reflect specified demographic characteristics of children attending school in the United States. Such demographic characteristics include socioeconomic status, ethnicity, region of the country, and size of school district. Percentile ranks are then derived from frequency distributions of individual students' scores at each grade. Norms provide a mechanism for describing a student's performance relative to that of other students in the same grade from across the country at a particular point in time.

The use of these norms to describe group performance must be interpreted carefully. For example, if a state's average score in reading is at the 54th percentile, the proper interpretation of this score is that the average, or typical, *student* in the state performed better than 54% of the norming sample. It is not appropriate to conclude that all students in the state are above average in reading, that the state as a whole is above average in reading relative to other states, or that the state as a whole is above average in reading relative to the national norm.

The approach used by some states and school districts in the reporting of group performance is to report the percentages of students scoring, say, "at or above the 50th percentile," or "in the average and above-average range." Although this method of reporting is appropriate because it maintains the relationship between individual performance and the national norms, the reported percentages should be accompanied by corresponding percentages for the national norming sample. Although it is obviously the case that 50% of the national sample of students scored at or above the national median at the time the test was standardized, it may not be the case that 50% of the national sample scored at or above the national mean raw score or national mean scaled score. If the reporting metric is something other than the percentage of students scoring at or above the national median, the appropriate national comparison should be provided so that proper inferences about the data can be made.

Interpreting Local Performance Relative to National Performance

It is unlikely that the demographic characteristics of the students in *any* state or school district mirror those of the nation as a whole; it is equally unlikely that the curriculum of any state or local district is as diverse as that of the nation as a whole. Furthermore, it is not necessarily the case that the guidelines set forth by the test publisher with regard to the testing of handicapped or limited-English-proficient students in a norming program are the same as those used in actual practice. If there were a state or district whose demographic characteristics matched those of the nation, whose curriculum was as diverse as that of the nation as a whole, and whose administration guidelines and procedures were consistent with those used by the publisher for the norming sample, one would expect the average student in the group to score at about the 50th percentile. To the extent that differences exist, we must remind ourselves that when local group summary scores are interpreted in reference to a national norm, the interpretation has to be placed in the proper context, simply that of the group's average student relative to the national norm. Because it is unlikely that the students tested in any given state or district are typical of the nation in all respects, it would be unreasonable to expect *any group* to be *at* the national average.

Test purchasers, districts as well as state agencies, often select tests through a process that examines the match between the test content and the local curriculum. In many cases, the selected test is the one that best reflects the local curriculum. Test users selecting tests on this basis may have an advantage over the norm group because the test is likely to be more valid for assessing performance in the local curriculum than it is for assessing the performance of a national sample of students being exposed to different curriculums, presumably having somewhat different emphases.

The Stability of Achievement Test Norms Over Time

Cannell's report suggests that the use of "old" norms is partially responsible for high achievement test scores. Presently, test publishers produce new editions of their tests on a 7-to-9-year cycle, and current norms are provided with each new edition. Because test adoption cycles do not necessarily coincide with test revision cycles, it is conceivable that the norms for a newly adopted test may be 2 or more years old. Therefore, it is critically important that empirical norming dates accompany the reporting of achievement test results.

It is very encouraging to note that today's students are performing better than their counterparts did in the late 1970s and early 1980s. Evidence of this improvement in performance can be found not only from research that test publishers have conducted in equating newly published tests to previous editions, but also from a recent research study conducted by The Psychological Corporation with the current edition of the *Stanford Achievement Test* Series. First standardized in the 1981–82 school year, the Stanford Series was administered to a nationally representative sample of 350,000 students in spring and fall 1986. The sample was further stratified according to "user" and "nonuser" groups, where "users" were defined as school districts that had been using the Stanford in one or more grades for at least one year in their districtwide or statewide assessments. The results of this study revealed that "users" outperformed "nonusers," and, more importantly, that "nonusers" performed better than the original norming sample in mathematics, reading, and the language arts in the elementary grades. Two important generalizations can be made from this research. First, test scores do tend to increase when the same test series is used year after year. However, this should not necessarily be attributed to "teach-

ing to the test"; rather, the test results provide a needed focus on areas in need of improvement. Second, educational achievement did improve from 1982 to 1986 in some subject areas in the elementary school grades. Therefore, more current norms for the Stanford Series have been developed and are available to users of the battery.

During this time of educational improvement, it is important not to lose sight of the fact that use of the same norms over a period of years enables the test user to demonstrate improvement relative to a constant reference group. Even if it were economically feasible for test publishers to produce representative national norms more often, frequently updated norms represent a "moving target," where educational gains (or losses) would be masked by the relative nature of the information. The level of achievement of students in the United States has increased in recent years, and educators must have the opportunity to demonstrate these gains in order to ensure the necessary support of the local community in improving the quality of education. The education of young people must continue to improve, and norm-referenced achievement tests are useful tools in this endeavor.

Conclusion

Because the public is expecting norm-referenced scores to represent standards of performance, we, as educators, must assist the public in becoming better informed about the interpretation of test results. National normative data provide extremely important information for making sound educational decisions. The degree to which these decisions are defensible depends on a clear understanding of the strengths and limitations of the data.

The Time-Bound Nature of Norms: Understandings and Misunderstandings

Paul L. Williams
CTB/McGraw-Hill

Presenting a view from CTB/McGraw-Hill, the author discusses the time-bound nature of test norms and argues that the phenomenon of most elementary students' scoring above averages from previous years' norms is a result of generally increasing levels of achievement.

Recent interest in the topic of the time-bound nature of normed scores has resulted, in part, from allegations made in a report issued by the Friends for Education. The key element of the argument put forth in the Friends for Education report is that too many students appear to exceed the national average. Data have been presented in the report which are said to show that more states and school districts are scoring above average than one might initially expect.

It is an interesting phenomenon that it is through the vehicle of the Friends for Education report that the time-bound nature of norms has received some measure of public attention. The fact that norms have always been referenced to the year of test standardization is something that has been so universally known and understood by testing professionals that it has not had a large measure of attention focused on it. Perhaps that will prove to be an important singular contribution of this issue of *Educational Measurement: Issues and Practice.*

The Cyclical Nature of Test and Norms Development

The evolution of norm-referenced tests (NRTs) as valuable assessment instruments has been characterized by the expansion of the purposes for testing. In the earlier versions of NRTs (in the mid-1960s to the mid-1970s), the primary purpose was to provide accurate normative scores so that group and individual comparisons could be made to a national profile of achievement. Using this information, school adminis-

Paul L. Williams is Director of Research and Measurement at CTB/McGraw-Hill, 2500 Garden Road, Monterey, CA 93940. He specializes in educational testing and measurement.

trators could evaluate programmatic and individual strengths and weaknesses so that appropriate instructional intervention and resource allocation could be applied. Additionally, using multiple-year testing, longitudinal trends in achievement could be monitored.

An expansion of these purposes took place with the publication of the *California Achievement Test* (CAT), *Forms C and D* (CTB/McGraw-Hill, 1977). This test battery, for the first time, allowed scores for instructional objectives to be reported from an NRT for individual examinees. Although earlier NRT test versions did allow test administrators to use item analyses for minimal diagnostic purposes, CAT C and D provided specific instructional objective scores for the purpose of more individualized instructional planning.

The schedule for the publication of norm-referenced tests has followed a basic, industrywide cycle of between 5 and 8 years for the same test series. In the instance where a test company has more than one NRT series, such as CAT and the *Comprehensive Test of Basic Skills* (CTBS), publication is staggered so that one test of the series is published about every 3 or 4 years.

This cycle has been dictated by several factors. The first factor has been the speed with which curricular changes take place in the nation's schools. NRTs are designed to reflect the predominant achievement outcomes and curricular trends in the nation's schools. When a new form of an NRT is developed, content considerations are of paramount importance. Although curricular trends have a major impact on the content of NRTs, these trends do not change so fast in the schools that more frequent revisions of a test series would be justified based solely on them.

At the time an NRT is revised, the collection of data for the generation of new national norms takes place. Using a national probability sample, data are collected for several hundred carefully selected school districts and hundreds of thousands of students. Based on this carefully selected stratified sample, normative scores are developed.

Each of the derived scores that emerge from the standardization process, including percentile ranks, grade equivalents, and normal curve equivalents (NCEs), has a predefined relationship to the characteristics of the norm group. Thus, at the time the test is normed, 50% of the examinees will exceed the 50th percentile and the same percentage will fall below the 50th percentile. Derived score tables for the test battery are produced, and all scoring of student tests is referenced to these tables until the battery is either revised or, in rare instances, when it is renormed with no change in the content of the test.

Data from national probability samples are not usually collected for a test more often than every 5 to 8 years because it is impractical and economically infeasible to do. It would not be reasonable to ask or expect schools to administer tests to large numbers of students every school year in order to develop yearly norms based on a national probability sample. The cost of such testing would have to be passed on by the publisher to the schools and would add substantially to the cost of school testing programs.

In summary, most large test publishers follow the common and decades-old industry practice of revising and standardizing their achievement tests about every 8 years. The content is updated to reflect current curricula and instructional practices, and new norms are developed so that the test reflects levels of achievement that prevail during the school year in which the test is standardized. The dates of standardization are given wide publicity, and all purchasers of the test are aware of these dates.

Proper Interpretations of National Norms

Because norm-referenced tests are not normed yearly on a national probability sample, changes in national achievement between the norming years will be reflected in the norm scores for groups of students. For example, if national achievement levels decrease between normings, as they did from the late 1960s to the mid-1970s, students' norm-referenced scores will decrease, and more students will fall below the median (50th percentile) score established when the test was normed. On the other hand, when national achievement levels increase between normings, more students will exceed the median established when the test was originally normed. Regardless of the direction of national achievement trends, when a test is renormed, exactly half of the students will fall above and half will fall below the newly established median.

At this time, national achievement indicators all point to the fact that student achievement is generally on the increase. This increase is documented by the National Assessment of Educational Progress (NAEP); the Scholastic Aptitude Test (SAT) results; two Congressional Budget Office reports (1986, 1987); and data collected during recent test normings by CTB/McGraw-Hill (1985, 1987, 1988).

Thus, during a time of increasing national achievement, the students' normed test scores will rise between norming periods. More students will score above the median score established during norming than will fall below it. This confirms the sensitivity of the test norms to changes in achievement, one of the tests' primary functions. These normed test scores are valid measures of student growth. Although the reference year for the scores will be prior to the year in which the test scores are reported, the test scores provide accurate program and student information. The fact that the norm scores themselves refer to norming that took place during an earlier year in no way compromises the major purposes for administering an NRT or the usefulness of the scores for program evaluation, student instructional planning, or the monitoring of longitudinal trends. When interpreting the scores, the test user must simply be aware of the year that the tests were normed and the general direction of national achievement trends. Interpretive guidelines are found in relevant test-related materials produced by most publishers.

The Friends for Education report has received attention primarily as a result of its improper interpretations of score distributions for

norm-referenced tests between renorming years. The sensational, and apparently illogical, phenomenon of having too many students above the national average is the basis for the criticism leveled at the testing community by the report. This is a point that should be elaborated upon, because it may be misunderstood by others as well.

A naive interpretation of what an average (mean) represents is that half of the scores in a distribution will fall above and half will fall below the average. Although this is a common interpretation, it is not statistically correct. The report suffers from this misunderstanding, as illustrated by the following quote: "Standard principles of mathematics make it difficult for more than one half of any group to be above average" (Cannell, 1987). There is no mathematical principle that would cause this to be so. Depending upon the shape of the distribution of scores and the measure of central tendency that is selected to describe the scores, more or fewer than half the scores may be above or below the measure of central tendency. For example, the mean, or arithmetic average, does not necessarily split a distribution of scores into equal halves. An average that splits the distribution evenly will occur only in a symmetrical distribution. If the distribution is skewed, there may be many more scores above or below the average depending upon whether the distribution is negatively or positively skewed. The median (the 50th percentile), on the other hand, does separate a score distribution into equal halves. Thus, there is no a priori reason to believe that norm-referenced scores should separate the examinees into two equal halves, particularly during times of changes in national achievement trends.

Extended Extrapolations

The time-bound nature of normative interpretations is relatively straightforward to describe and understand. What becomes more difficult to evaluate are the social and educational implications that might be drawn from acknowledging that actual score distributions may differ increasingly from the published norms as a result of changes in achievement over time.

One way to determine the amount of change in achievement over time might be to survey states and school districts and, based on the aggregation of scores, determine the number of states and districts reporting above "average" (50th percentile) scores. Additionally, it might be possible to determine the proportion of students above the 50th percentile and the average national student score. Finally, to illustrate the rapidity with which standardization score distributions change, data could be collected the first year after norming and then an average of state and district scores could be calculated.

This task would be very difficult to do correctly. Different states and districts use different tests that are not on a common scale. The scores from all states and districts would have to be collected, placed on a common scale, and analyzed appropriately. There is no evidence that this has ever been done correctly.

This brute-force approach need not be the only mechanism to determine achievement trends over time, nor is it the best way. Achievement changes between normings are documented by the major publishers, and this information could be directly examined.

A third approach intended to monitor national achievement trends might be NAEP. But NAEP is also an imperfect panacea for determining achievement growth. There will always be quality-control issues, as evidenced by questions about recent NAEP survey results. NAEP is a valuable indicator of achievement trends, but like any method it is not absolutely perfect.

The fact is that various sources of information must be synthesized so that a complete picture of national trends can be obtained. Each type of assessment, via NRTs, CRTs, NAEP, or others, attempts to answer different questions in different ways. Each is valuable in providing a piece of the picture on the status of student learning. It is when we learn how to make artful syntheses that all of us will be closer to determining the status of achievement in America's schools.

It is unfortunate that during a time when national achievement trends are moving upward some might use that fact to suggest that one of the reasons for the upward movement is inadequate norming by test publishers and inappropriate teaching of test content by users for self-serving purposes. These are serious charges that should not be made without supporting evidence.

It must be stated that there is no logical reason why test publishers would wish to engage in inadequate norming. Test publishers have every incentive to make sure that their tests are completely objective and are administered properly and that their integrity as valid measures of performance stands unimpeached. Without such quality, test publishers would quickly find themselves with no customers.

Conclusions

To be sure, some of the concerns raised by Dr. Cannell are shared by all in the educational community. The time-bound nature of norms may not be well understood by some school personnel and the public. There may be abuses of tests and breaches of security. Some teachers and administrators may indeed disclose too much test content to the students. But the overwhelming majority of the educational community is doing its very best to administer tests and report test scores in a responsible fashion.

At least two examples of this come prominently to mind. The first is the way in which test publishers equate alternate forms within the same test battery over time. Thus, CAT Forms C and D (1977) were equated to CAT E and F (1985). Similarly, equating is done between different test batteries developed by the same test publisher, as was the case for CTBS Forms U and V (1981) and CAT E and F (1985). These equatings allow the test user to move from one version of a test to another and preserve longitudinal comparisons. The recent trend that has been observed in these equatings is that the derived scores from the most recently normed test are lower than for the earlier normed test. This is predictable in times of increasing national performance. The opposite would be true if national achievement trends were on the decrease. Explanatory material that helps the practitioner understand this phenomenon is

always provided.

The second example relates to the Annual National Normative Trend Data (NTD) published by CTB/McGraw-Hill. Research on this project began in 1984, when an emerging customer need was identified by the company. Customer comments about the desirability of obtaining more recent normative data were noted in market research efforts. Such data could be used to amplify the standardization norms and provide a more complete picture on the progress local school districts were making in their instructional efforts. After 3 years of research, the NTD service was offered to CTB customers. Score reports have been made available on an annual basis, for the standardization year as well as for the most recent norming. This service is a response to those educators who have been concerned about the time-bound nature of norm-referenced scores.

The test companies do their best, through many vehicles, to assist the test consumer in being a responsible user of test results. Indeed, reasonable testing programs, effectively implemented, are one of the reasons that achievement is increasing and that we are not currently in the decline phase that manifested itself in the late 1960s to the mid-1970s.

The assertion that scores are on the increase does have merit. Perhaps the positive side of this phenomenon should be stressed more. States and local school districts have committed considerable resources to improving the achievement levels of their students. All indicators of student achievement appear to converge on this fact, particularly for the elementary grades. The American public should be gratified that achievement is increasing.

Cannell (1987) charges that "inaccurate initial norms and teaching the test," rather than improved achievement, are reasons for improving scores on nationally normed tests. The problem with these allegations is that there is little, if any, evidence to support them. To the contrary, the body of independent evidence suggests that test norms provide a valid and useful reference in both the norming year and in subsequent years and that achievement at the elementary level has been increasing. If indeed there exist instances of abuse of test norms and of misunderstanding of their meaning by educators or the public in general, then the proper remedy should be to correct those instances rather than to make rash allegations about the adequacy of test norms or questionable teaching by educators.

References

Cannell, J. J. (1987). *Nationally normed elementary achievement testing in America's public schools: How all fifty states are above the national average.* Daniels, WV: Friends for Education.

Congressional Budget Office. (1986). *Trends in educational achievement.* Washington, DC: Author.

Congressional Budget Office. (1987). *Educational Achievement: Explanations and implications of recent trends.* Washington, DC: Author.

CTB/McGraw-Hill. (1977). *California Achievement Test, Forms C and D.* Monterey, CA: Author.

CTB/McGraw-Hill. (1981) *Comprehensive Test of Basic Skills, Forms U and V.* Monterey, CA: Author.

CTB/McGraw-Hill. (1985). *California Achievement Test, Forms E and F.* Monterey, CA: Author.

CTB/McGraw-Hill. (1987). *Annual national normative trend data.* Monterey, CA: Author.

CTB/McGraw-Hill. (1988). *Annual national normative trend data.* Monterey, CA: Author.

SRA Response to Cannell's Article

Audrey L. Qualls-Payne

Science Research Associates

The author defends SRA's norms, discusses some of the difficulties in pursuing Dr. Cannell's proposals, and points out that we need to monitor not just student achievement levels but also trends in curriculum.

Science Research Associates (SRA) recognizes the concerns expressed in John Cannell's article, "Nationally Normed Achievement Testing in America's Public Schools: How All 50 States Are Above the National Average." We differ, however, in our assessment of the situation and the proposed alternatives. According to the article, most schools in the nation perform at or above average on commercially available tests. This finding, as noted by Dr. Cannell, is not consistent with statistical theory, which says that half the students should be above and half below. Dr. Cannell expresses the opinion that this inconsistent statistical phenomenon results from using older tests, older norms, teaching to the test, statistical manipulation of the data by publishers, excluding special

Audrey L. Qualls-Payne, Psychometrician, Science Research Associates, Inc., 155 N. Wacker Dr., Chicago, IL 60606-1780. She specializes in achievement testing.

BEST COPY AVAILABLE

education students from the calculation of group averages, and inaccurate norms. He goes on to suggest that these problems can be eliminated by the use of one achievement test in all schools across the country with the concurrent development of annual norms. Our purpose is to examine Dr. Cannell's conclusions and offer alternatives to some of the issues raised in his report.

SRA's national norms are reliable and accurate indicators of national student performance at the time of standardization. The charge of statistical manipulation of data appears to result from Dr. Cannell's apparent misunderstanding of the purpose of the various types of test scores and subgroup norms. Schools may wish to compare their students' performance with, in addition to that of the national group, that of groups more similar in structure and student composition. For example, a nonpublic school may want to compare their students' performance with that of students from other nonpublic schools. The various test scores, in addition to status scores (i.e., percentiles and stanines), are offered to meet the many needs of our customers. Normal curve equivalents (NCEs) are required for Chapter 1 program evaluation. To assess longitudinal growth and determine functional levels, developmental scores, for example, standard scores and grade equivalents, are needed.

Dr. Cannell's alternative to the various standardized achievement tests is a national achievement test, which would require at least two major actions. First, this national achievement test would have to be normed annually with a representative group of students to have yearly norms. Second, new test forms would be needed for each administration to eliminate possible problems of teaching to the test and test security.

A project of this magnitude and complexity would be very difficult logistically and very costly. Two major logistic problems would be (a) obtaining curricular consensus on the test content and (b) obtaining or mandating national participation.

If yearly new forms are not an option but annual norming is, and if there truly is a substantial amount of teaching to the test, the problems noted in Dr. Cannell's analysis may not go away. If new forms of achievement tests are developed each year, thereby increasing test security, the need for annual norms diminishes significantly. Based on Dr. Cannell's analysis from schools with tight test security and literature on student growth, drastic shifts in student performance from one year to the next are rare. From a psychometric point of view, new norms are needed only when there is a *significant* shift in school curriculum and/or student performance.

As opposed to developing and standardizing new forms each year, a mechanism is needed to monitor changes in school curriculum and student performance. Whenever there is a change in either curriculum emphasis or achievement levels, new test forms should be developed and standardized. If the change is strictly a shift in student achievement, renorming is required. As a publisher, we must base our decision on when to issue new forms/new norms on a systematic monitoring system.

There are several ways to monitor student progress. One way to accurately spot when significant changes are taking place is to track student achievement on a regular basis (i.e., annually). The entire user group could be used for this purpose. The monitoring process should be capable of producing user-based norms, which can then be made available to all customers as an optional service in addition to the national norms.

There is at least one major problem with the user-based monitoring system. If the user sample is biased and unrepresentative of the national student population, significant changes noted in the user sample may not truly reflect changes at the national level. One way to resolve this problem would be to select a subset of schools from the user group and use it to monitor changes in curriculum and student achievement annually. The selected schools should be representative of the national population of schools with respect to geographic region and racial/ethnic and socioeconomic status. Once a set of schools is selected for this purpose, students in these schools can be tested on an annual basis and norms can be developed. As in the previous method, annual norms will be made available to customers as an optional service. Because of the representativeness of the schools selected for monitoring, one can, with a high degree of confidence, generalize results from this set of schools to the U.S. population of schools.

Because SRA recognized the value of a monitoring system, we are already in the developmental stages of implementing such a program.

Senator PELL. I think, Mr. Tucker, earlier you said that the utility of these tests for improving education is really determined by how those tests are followed-up in the schools. Could you enlarge on that thought?

Mr. TUCKER. I'm sorry, Senator.

Senator PELL. I think you said the effectiveness of a testing program is determined, mostly by following up on the results.

Mr. TUCKER. Oh, absolutely. That is to say, it's the point Al made about consequences, I believe is what you're talking about. That is, what you find in Europe, for example, and let me just take Germany as an example, you can get an apprenticeship as an automobile mechanic with the local garages or you can get one with Mercedes-Benz at Daimler Benz. Whether you get the one at Daimler depends upon the grades you've gotten, the courses you have taken, and the exams, the exam grades that you have gotten.

So, for the whole system, whether you want to be an automobile mechanic or whether you want to go to the finest university in Germany, how you do on the exams makes all the difference. And that is the whole point of the examination system. It is the consequences that go with it.

That is why kids in Germany, whether they want to go to a fine university or whether they want to go to Daimler Benz, take math, tough math, tough science, and work very hard in those courses. And that's why ours don't, because for 90 percent of the jobs or more that you get coming right out of school, it doesn't matter. In fact, if you take a tough course, what you risk in the United States is you get a nonpassing grade.

Senator PELL. Thank you very much.

Am I correct in saying that my good friend, Ira Magaziner, is the chairman of your board?

Mr. TUCKER. No. He is the chairman of the commission that we put together last year to produce "America's Choice." That, "America's Choice," was produced by the Commission on the Skills of the American Workforce, which we put together, and it's one of four programs in our organization.

Senator PELL. Thank you.

I will not ask any more questions at this time.

Senator Kassebaum.

Senator KASSEBAUM. Thank you, Mr. Chairman.

I suppose because I agree totally with Ms. Daniels' statement, I happen to believe it's a very strong statement, but I think you make some very good points. Just to comment on one, you mentioned parents as a part of the equation. As a teacher, I am sure you have found many parents have come to you and said, "I don't know why my son or daughter is not getting an A or a B," when in all fairness to that student they should be getting a C or a D.

So while parents on the one hand want a good educational system and want their children to be doing well in education, we haven't yet gotten to a point where they are willing to see their child get a C or D and spend the energy on what it takes to raise the level.

Sometimes, I suppose, for teachers it's just easier to say, "Oh, well, give them an A or a B."

Ms. DANIELS. It really depends on how much you want to tough that kind of thing out, because it is a constant dilemma. And in one school that I taught, a veteran teacher there took me aside and said, "Maureen, I just want you to know that the parents in this community expect their children to get good grades."

Now, I want to impress upon you that they don't make the connection that the children should be earning those grades necessarily. They expect you to give them those grades. And, unfortunately, that was the mentality among many of the parents. And you're absolutely right, Senator, a lot of parents, because so many of our mothers work outside the home now, which wasn't the case when you and I were in school, that further exacerbates that situation, that parents are not there to reinforce and do what needs to be done, you know, the magic of the success of a student is what we do in school but almost more importantly what they come to us with and then what's reinforced once we send them back home. And we constantly fight that now.

Senator KASSEBAUM. That is why I am not sure I think testing or even a certificate of mastery helps that situation until we begin to change our thinking about what we are willing to invest in it as parents and as participants in the community.

Mr. TUCKER. I think, from my point of view, Senator, that there is a couple of points to be made here. One is, grades don't mean anything in school with respect to any specific standard. We did a big survey in Rochester, NY, last year, and we interviewed, among lots of other people, kids. Poor, black, minority kids in Rochester told us in overwhelming numbers they get put on honor roll—I am quoting the kids—"And all I did last year was color in the maps."

Now, why does that happen? It happens because an estimate is made of their ability in grade 2, 3, or 4 or maybe late as grade 5, that they aren't very smart, they aren't very capable, and they aren't going to make it. And so we are going to give these kids very undemanding courses, very undemanding material, and we are going to give them good grades because they tried hard. OK.

The record shows this happens all over the United States, and the kids don't find out until later.

Senator KASSEBAUM. I bet Ms. Daniels doesn't do that in her classroom.

Mr. TUCKER. No, I am sure she doesn't. But the record shows that this happens in very large numbers all over the United States.

Senator KASSEBAUM. Let me just ask, Mr. Tucker, shouldn't it be a school board who would be demanding a different approach?

Mr. TUCKER. When was the last time they did that, Senator?

Senator KASSEBAUM. Well, I was a school board member. I know it isn't easy.

Mr. TUCKER. No. I was too, and the dynamics of the school board that I was on and many others that I have observed don't lead you there because—well, how do I put this?

Senator KASSEBAUM. Well, no, I understand what you're saying.

Mr. TUCKER. Right. I mean, all the pressures are the ones you're reporting on. All the pressures are to give kids good grades so that they can get the rewards that grades produce in our society. And the problem is they are unconnected to achievement.

As an entire country, we are fooling ourselves. The record showed in the last big survey that was done of us vis-a-vis other countries' performance in math and science, we performed worse, and when our kids and parents were asked how we did, we said we did terrific. The kids who were performing way above ours, they said not good enough, not very good.

Mr. FALDET. Senator, if I may. You make a good point in that the test scores can at points be supportive of teacher judgment. But test scores are frequently viewed as Caesar's messenger. If you don't bring good news, you may get shot.

Senator KASSEBAUM. Oh, yes. That's right.

Mr. FALDET. And that is one of the problems and, I think, has resulted in a lot of the "Lake Woebegon" effect and a lot of the criticisms at tests when it's the message that you need to pay attention to, not the messenger. Thank you.

Senator KASSEBAUM. Thank you.

Thank you very much.

Senator PELL. Thank you very much.

Senator Bingaman.

Senator BINGAMAN. Thank you, Mr. Chairman.

Let me ask a sort of general question for Mr. Tucker and Ms. Daniels, and any of the rest of you who have ideas about it. I have the idea that perhaps some type of standardized national testing would help us, would help teachers and school boards and others who are interested in improving education in reforming the educational system, to actually get more attention to the courses that matter, because I fear that what we have had over the last several decades—and I don't know if this is accurate; you folks who are experts could tell me—but I fear that we kept the number of hours of instruction fair.y constant in our school system.

In some States it has increased slightly, but it hasn't increased significantly. We have loaded the schools down with more and more responsibilities to do driver education, to teach nutrition, to teach sex education, all the things that schools didn't used to do. And the basic courses that we are adopting national goals in are not those. We are saying we want to be first in the world in math and science, and yet we don't have the number of hours in the week committed to math and science instruction so that we could begin to approach accomplishing that goal.

And if you had a national test which someone could call up and say, "Now, look, we are going to be tested on math, and we've got to have these 6th graders who have got to learn math regardless of what else falls off the table," I mean it begins to build some other pressures into the system, which I don't think are there now.

I guess, Mr. Tucker, if you have a thought on that?

Mr. TUCKER. Well, I mean, I would say that is an eloquent expression of my point of view, yes. I mean, I think if we don't know where we're going, as the old saw says, we won't get there. And there is nothing that focuses the mind quite so nicely as an examination—I want to come back to Lauren's point—an examination in which it is clear what the kids are supposed to know and be able to do. An examination in which that is not clear, which is actually true of most American tests, won't produce the result you're talking about. But if it's quite clear what kids are supposed to know

and be able to do and if in fact it is clear that that statement is the outgrowth of a real national consensus, then, yes, I believe it will have a powerful effect on what goes on in schools.

I would like to say, however, on the question of whether that means a national curriculum, my own view is that such a test need not lead to such a curriculum. I think it might actually do the reverse, in the following sense: Most observers who have looked at America's schools from other countries are stunned at how similar our schools are all over the United States, what is taught and how it's taught, much more so actually than in countries that have examinations.

My own view is that we ought to set up a system in this country in which it is very clear what kids ought to know and be able to do. We ought to do that at a national level. We ought to have one or more national examinations which are all keyed to that same set of standards. And then we ought to tell people like Ms. Daniels, our teachers in the school, "You are free, you are free to figure out how to get the kids there and get rid of most of the State rules and local rules and regulations that affect the freedom of the teachers to decide how to teach these kids."

We haven't done that up till now because it has never been very clear where we were supposed to go. So in fact we now have tons of rules about what you're supposed to teach and how you're supposed to teach it. We could get rid of most of those rules if in fact we were clear about where the kids were supposed to go.

Ms. DANIELS. I think teachers have a very clear sense of where children ought to be and where they need to go, and our hands are tied, in most instances, to get there.

I am disturbed, in a sense, by the focus on preparing students for math and science. The goal that speaks to me is that we need to make sure all our children are prepared when we come to school, and we don't seem to be focused on that very much.

Senator Wellstone spoke to that earlier. But I can guarantee you, unless we address goal 1 and do something more than pay lip service to that goal, that we can pretty much predict the rest of the goals aren't going to fall into line. That, to me, is the foundation for achieving all the rest.

So, you know, we know some things. We know poverty is a valid predictor of school failure. We know that children that are malnourished and neglected, that are homeless, those are children that we are dealing with now. We know that they come into the school system disadvantaged, and those kinds of problems are special challenges on teachers right now.

So we can get to the math and the science, but we need to address the basic human needs of some of our students before we can educate them. And that is what I would like to see the discussion center on, and we seem to be way far ahead of where we need to be at this point.

Senator BINGAMAN. Well, I don't disagree with you that the basic needs have to be addressed before we can accomplish the rest. But I think that, like everything in life, if you wait to do something until something else is done, you never get the second thing done. And I tend to think you have to work on different issues in parallel

and make as much progress as you can on various things at once. So I guess that is where I am coming down.

Let me ask if any of the witnesses, and, Mr. Tucker, you said you do favor a national test, do you think it should be a mandatory national test, or should it be voluntary?

Mr. TUCKER. I do not think that the Congress, certainly at this time, should mandate a test. I think actually that, given the enormous fear in this country of anything that looks like a federally imposed curriculum, that could actually set back considerably a move toward a national examination system.

My own view is that we probably will move eventually toward a system in which it will be mandatory for all kids to take some form of examination which can be calibrated to a standard examination. That is what we have proposed.

And I think that would be a good idea for the country, but I think there is no evidence that if Congress tried to mandate that now we would move there sooner than if we did it essentially through the normal course of events.

Coming back to what I said a moment ago, I think what is really important now is to start moving toward agreement, and a number of other witnesses have said this, on what the framework for the standards ought to be. The Roemer panel is moving in that direction. I believe that the administration is very sympathetic toward that.

I think some form of either the Roemer panel or what you have proposed at various times can be a vehicle for moving us toward agreement on a framework, and that is probably what we should be concentrating on now, plus the other two things I mentioned, which is support for people doing research on new forms of assessments. We aren't going to produce a good assessment unless people are working to produce it.

Third, and I just feel terribly strongly about this, we have got to make an investment in figuring out how to bring kids up to a high standard because the exam itself is not going to do that. The exam will be a very powerful tool for giving incentives to lots of people to try, but if you don't put in place, start putting in place now, the mechanisms that are going to lead to much more productive school systems than we've got now, the exam is not going to do the job.

Those are the three things I would have the Congress do now.

Senator BINGAMAN. Thank you very much, Mr. Chairman.

Mr. FALDET. Mr. Chairman, I beg your indulgence. Mr. Melody is accompanying me, and would it be possible for him to end with just a very brief statement to the committee?

Senator PELL. Yes. Certainly.

Mr. FALDET. We would appreciate that.

Mr. MELODY. Thank you, Mr. Chairman, members of the committee. I am here to second something that Burt said a few moments ago. And that is that we in educational publishing believe we are part of the process, and we are as committed as are you to working toward the resolution of some of these problems. I am a senior executive of one of the Nation's major text and test publishers, and I am simply here to demonstrate the sort of commitment that both my company and the industry has for the resolution of our problems.

Senator PELL. Thank you very much indeed.

I thank all of you for being with us. The record will stay open for 2 weeks for the insertion of any further statements that are made.

[Additional statements and material submitted for the record follows:]

PREPARED STATEMENT OF GOVERNOR ROMER

Mr. Chairman, members of the committee, thank you for the opportunity to submit testimony to the Subcommittee on Education, Arts, and Humanities of the U.S. Senate. The topics we are discussing, the pros and cons of national education testing, educational assessment, and education reform are vital ones. I regret not having been present to participate in your hearing with such distinguished colleagues as yourselves and the panelists.

I have the privilege to serve as the co-lead governor for education for the National Governors' Association and as chairman of the National Education Goals Panel. I want to make clear, however, that I am here today in my capacity as Governor of Colorado and am not speaking in any formal way for NGA or the goals panel. The ideas I will discuss have been, and will continue to be debated by the governors and by the goals panel. They are creating interest and excitement, but neither the National Education Goals Panel nor the NGA have yet adopted a formal policy on the issue of national education testing. I want to begin by giving you a very brief background on the National Education Goals Panel. National Education Goals were established to provide a common framework and vision for education reform. As governors, we wanted to be clear that we were committing ourselves to making education a priority in each of our states and were willing to be held accountable for this commitment.

The goals themselves have been criticized as too ambitious. In my judgment, they accurately identify what our reach must be to make the kind of educational progress that is required to secure our future. The goals are not gospel. They are not perfect. But they do provide a common vision, a common way to organize reform issues, and a common language.

Last July, the National Governors' Association adopted a policy establishing the National Education Goals Panel to oversee the development and implementation of a national education progress reporting system. Beginning in September 1991, the panel will issue an annual progress report to the Nation on our progress toward achieving the national education goals.

The panel's work involves two tracks: One short-term and one long-term track. The short term agenda is to produce a credible report to be released in September 1991. In working toward this deadline, the panel is aware that great harm can be done by asking the wrong questions or by trying to force existing data into new forms. Accordingly, the panel is committed to a thoughtful and comprehensive approach over the long term, even if it means leaving some blanks in the first report.

Our long-term agenda involves encouraging the development of an assessment system that will not only measure our progress toward the goals, but that will help us restructure the education system to achieve the goals.

In an effort to take advantage of the expertise that exists in the country on these issues, the panel established six working groups to give advice and policy options to the panel about its work, both short-term and long-term. These groups are in the process of developing creative approaches for measuring our progress in each of the six goal areas. A list of those working groups is attached to this testimony. Their recommendations to the panel will be released late this month.

With respect to goal No. 3, increasing student achievement, the panel has identified three waves of action and effort that will need to be mobilized if we are to make substantial progress toward the goals by the year 2000:

—Defining specific standards of student achievement (what a student should know and be able to do);
—Designing an assessment system that will allow us to accurately and authentically measure our performance against these standards; and finally,
—Moving the education system to close the gaps between current levels of performance and the standards.

These tasks lie along a continuum from national to local effort. In the panel's judgment, national goals and national standards for student performance are appropriate and necessary. At the other end of the continuum, we believe that the design and implementation of specific reform strategies must remain locally based in the

best of our traditions. In the middle, the assessment challenge provides an opportunity for joint action. Let me discuss each of these in turn.

Standards. We need to begin with detailed goals and objectives for learning. These objectives must be uniformly high and must reflect what future citizens of our Nation will need to know and be able to do. Given the diversity of reform strategies, an overall national educational standards framework is needed to provide the targets at which all would aim their efforts.

Discussion regarding the process by which these standards can be set has just begun. The following elements have been identified as critical to this process:

—There must be a key role for governors and the states, which still have primary policymaking and funding responsibility for education.

—The process must be highly consensual, participatory and inclusive Broad consultation with all segments of our population is the only way to build the commitment of all Americans to the hard job of educational change. In fact, the standard-setting process can itself become part of the renewal process if it represents a shared clarion call for higher expectations and higher standards for all students.

—To make sure the standards we set are competitive, we will need to use as a benchmark the best standards in the world.

—The process would set standards, not prescribe how the standards are to be achieved. Common standards need not and should not give rise to a national curriculum that dictates to teachers what and how they must teach or that limits the choices of school districts or states in selection of textbooks or instructional materials. The standards would be a description of the concepts, knowledge, and skills that students should master, not a recipe for achieving those standards.

Assessment. Tests should have the capacity to support reform and not just measure present performance. Tests also should have the capacity to measure the skills and abilities students will need to know.

I am concerned that many of the assessments we use now don't authentically measure higher order thinking skills. We rely on standardized tests that tend to drive curriculum and instruction toward rote memorization and the discovery of a single "right" answer. I am also concerned that the tests on which we rely to provide national and state samples of performance don't give the kind of feedback individual students need to help them and their parents judge the adequacy of their performance against national standards. These tests are useful and I support their continuation, but for assessment to stimulate better performance by students, the test ultimately has to be important to each student.

The notion of testing all students rather than a sample of students raises legitimate concerns about local control of education. There is a fundamental tension between the compelling need for national goals and standards and the tradition and strength of local control. We know that the most effective schools are those that are managed at the school level. We draw strength from the diversity of approaches that local control makes possible.

Therefore, I do not advocate a single Federal test. Instead, I am interested in exploring the alternative of a national examination system. This system would reflect not only the voluntary participation of states, but the powerful combination of their individual expertise. The system would be composed of several examinations anchored by a common set of national standards.

States, which already devote enormous resources to testing, could work in clusters to develop and implement examinations. Each cluster's examination would be designed to assess the range of knowledge and skills encompassed by agreed upon national performance standards.

Clusters might be formed around different interests and states would be free to decide which cluster examination was most appropriate for their use in any given subject. The interests of clusters could evolve over time and states would be free to change their associations.

Through an anchoring or calibration procedure, a means for equating student performance across clusters could be devised. The key to this process is a commonly agreed upon definition of what students should know and be able to do. These examinations cannot, strictly speaking, be designed until national standards have been developed and adopted. However, we can begin work on the process of designing and evaluating examinations formats. We can develop a format which will work will in the anchoring system and begin to experiment with different methods of calibration.

In suggesting this approach, I do not wish to minimize the degree of technical work that will be called for nor the time required for developing the system. However, at this point, it appears that this type of assessment system is both technically feasible, and essential to meeting our national reform objectives while retaining, and even strengthening, our tradition of local control.

Motivating Changes in the System. The changes that this type of assessment system could help drive include:

—Schools will offer challenging learning opportunities for all students, not just those who are college-bound. For most of the 20th Century, America has had two systems of public education—one designed to train a small group for management and leadership and another to prepare the remainder for the routine work they would do as adults. We now recognize that a workforce dominated by individuals who are not challenged to achieve higher than 8th grade skills is uncompetitive in a greatly changed—and changing world economy.

—Learning environments will be structured to encourage and reward student effort.

—Parents will have the knowledge and motivation to be effective partners with the school because they will understand the skills and abilities their children must have to prosper economically and how their children are performing in relation to this standard.

—The means by which we assess performance will accurately track what we want students to know and be able to do. There is a consensus emerging that the new basic skills include critical thinking, information management, interpersonal and communication skills, and problem solving. Not all classrooms and very few tests presently focus on these skills.

—Assessment of student learning will have consequences. If employers and colleges participate in setting the standards on which the examinations are based, and agree to take performance on the examinations into account when deciding who will be admitted to college and who will get hired at what pay level, then millions of students will, for the first time, see a direct connection between their performance in school and their opportunities in life.

I am a pilot. The system I experienced in obtaining my pilot's license provides a useful model of the kind of system we are aiming toward. When I began my course of study, I clearly know the competencies and understandings that I would have to demonstrate to obtain the license. And I directed my effort at achieving these understandings and competencies. Some came faster and easier than others. The length of my course of study did not depend on an arbitrary number of seat hours, but on mastery of the required skills. I established my command over the required subject flatter through a demonstration. The test therefore related directly to the subject matter I was being taught—and to the skills I need to pilot an aircraft safely.

An examination like this one, that can organize and motivate effort, that can accommodate individual differences in learning rates and styles, that provides real consequences for the learner and that establishes standards of performance that all students are expected to meet has much to offer as we reflect on the role assessment can play in education reform.

I welcome a continuing debate on these issues. As chair of the National Education Goals Panel and co-lead governor for education for NGA, I hope and intend to have a close working relationship with this committee.

Thank you again for the opportunity to participate in this hearing. I would welcome any questions or comments from members of the subcommittee.

Prepared Statement of Mr. Melody

Mr. Chairman and members of the committee, my name is Michael E. Melody. I am senior vice president for College and Test Publishing for Houghton Mifflin Company. Houghton Mifflin Company is a publicly held, Massachusetts corporation which has been located in Boston for over 150 years. The company is a major publisher of textbooks and other educational materials, including assessment instruments, for schools and colleges, general interest and reference books for adults and young readers, and computer software for educational and business applications, for both domestic and international markets. I welcome this opportunity to contribute to your important work.

We employ over 1,400 people in Massachusetts and about that many more throughout the rest of the country. Houghton's annual Massachusetts payroll is

more than $45 million, and we spend over $125 million locally with freelance writers, artists and photographers, printers, paper merchants, trucking companies, and other suppliers of goods and services each year.

Among our employees located elsewhere in the U.S. are nearly 400 people at two locations in Illinois—our midwestern regional office in Geneva and The Riverside Publishing Company in Chicago.

The Riverside Publishing Company, a wholly owned subsidiary of Houghton Mifflin, publishes standardized tests in a variety of formats. Perhaps our best known program is the Iowa Test of Basic Skills which was first developed in 1935 under the direction of the staff of the College of Education at the University of Iowa in Iowa City.

Houghton Mifflin first published the Iowa tests in 1940, thereby making them available to schools nationwide. Our relationship with the University of Iowa, now through our subsidiary The Riverside Publishing Company, is one which has flourished for over 50 years. Millions of dollars and volumes of research data have been invested by both the University of Iowa and the publisher to developing and revising this testing program in order to keep it up to date in order to meet market requirements.

In addition, Riverside publishes the Integrated Literature and Language Arts Portfolio Program, a performance-based approach to evaluate student accomplishments in these critical areas of instruction and is developing a state program in performance-based testing of reading, language arts, and mathematics for Arizona. This project has attracted nationwide attention.

My primary purpose is to use this statement to provide a context for the consideration of the pros and cons of a national assessment program, from the experience base of a commercial publisher of standardized, multiple-choice tests. This is a particularly relevant perspective inasmuch as one rationale for a national assessment program is to compensate for the alleged deficiencies of such tests.

Houghton Mifflin and The Riverside Publishing Companies, and the other members of the AAP, remain committed to providing educational assessment instruments that are fair, valid, and reliable; to making use of technology; and to providing assessments with diverse formats, including multiple-choice, performance assessments and portfolio programs.

As Congress and others look at the issues involved with a new national assessment program, an understanding of current testing practices and the uses and limitations of testing should be helpful. Test publishers provide one—but only one—of the key elements in the instructional process. Our instruments provide information on individual performance, based on a sampling of skills and knowledge. Contrary to what is becoming the conventional wisdom, our tests can assess both basic skills and higher order skills, and can be administered in a variety of formats, such as multiple-choice, where the pupil selects an answer that is either right or wrong, or a "performance" format, where the student's performance is considered to be better or worse in terms of how that answer is arrived at and delivered.

The problems of education in America today can not be attributed simplistically to inadequate or inaccurate information derived from standardized tests. Standardized tests are not the only source of information on how well a student or a school system is doing. We have not made that claim. To do so would overstate the power of a test and undervalue the role of the teacher and the school system to provide relevant and accurate information.

Test-derived information can be used for many purposes. We believe that the most important use is in diagnosing individual strengths and weaknesses to improve the instruction of that pupil. Scores can also be used to determine how an individual compares to students nationwide as well as whether that individual has mastered local, State, or national instructional objectives. Individual scores can be used by the local and state education authorities for their own accountability purposes, to evaluate the results of the effectiveness of instruction, and to evaluate and improve curriculum.

To assure that tests are used effectively and appropriately, publishers provide specific guidelines and suggestions for the interpretation of test scores and how to present the scores in ways that both pupils and parents will find helpful; how to diagnose specific strengths and weaknesses; and how to use test results to improve classroom instruction. We provide in-service support to teachers and test administrators. These efforts are in accordance with the Standards for Educational and Psychological Testing and the Code of Fair Testing Practices in Education. (Copy attached at end of this statement)

To assure the school systems, parents, and the public of the quality of the tests being used, publishers also document the reliability and validity of their tests, and

support their tests through extensive research and development efforts. Riverside's test authors are eminent educators and psychometricians with expertise not only in test construction, but also in curriculum development, instructional application, and the psychological implications of testing in school and university settings. Over 50 years of research at the University of Iowa and over 300 research studies have produced information about test construction that is the foundation of excellence on which Riverside's tests are built.

These same standards that have been applied to multiple-choice tests are also appropriate for performance assessments. It is important to stress that these formats are complementary. Performance assessments should not be considered a panacea, nor as a mutually-exclusive alternative to multiple-choice tests—each serves special purposes and has unique advantages and disadvantages.

Performance assessments are not new. Teachers have traditionally evaluated writing skills by grading papers that students have written. Riverside's Iowa Tests of Basic Skills and Tests of Achievement and Proficiency include writing supplements for grades 3-12 that have standardized this process by giving students a standard set of topics and standard guidelines for completing the essays. The essays are then scored using a standard procedure. Since all students follow similar directions and are graded in a consistent manner, the score for each writer's essay can be compared to national norms.

As mentioned earlier, Riverside also is developing a statewide performance assessment for Arizona. It will complement the Iowa Tests of Basic Skills, enabling assessment of progress over the school year from both a national and local perspective. The new test will be administered to grades 3, 8, and 12 in the spring semester. It requires written responses to questions, including short answer and essay responses in the reading and writing sections, and graph and chart constructions in the math section. Arizona teachers will be trained to score the new tests, to assure that the results will be fair and comparable. Dr. Monty Neill of FairTest spoke positively of this project in his comments.

While it is true that performance assessments are not new, it is also important to point out that the use of this format for a wide range of subject areas in a high stakes context, where the results are to be used for comparisons of individual students and for system accountability, is breaking new ground. Accordingly, these expanded uses for performance assessments should be approached very cautiously to assure that they do meet the essential requirements of fairness, validity, and reliability.

Which format is used, of course, begs the question of what is being assessed. It is unfair to criticize multiple choice tests for not providing information on how a pupil is performing on national standards when those standards do not yet exist. In the absence of a national consensus on a standard for what should be taught and tested, standardized tests that compare a pupil to students nationwide have been the only objective and accurate appraisal of achievement beyond the local curriculum. We are now in a transition period, during which a consensus is growing that minimum national educational standards are appropriate and necessary. But that consensus is not yet attained nor has it been tested.

Is there a national curriculum now because publishers sell their texts and tests nationwide? We believe not. There still remains considerable diversity around the country on curriculum and the scope and sequence of instruction. Recent advances in electronic publishing also make it far easier and more cost-effective for us to accommodate the demands of our customers to respond to their needs. It also makes it easier for us to respond to rapidly changing national educational movements and the resultant market demands.

To provide one example, in 1988, the Bradley Commission on History in Schools released its report, "Building a History Curriculum: Guidelines for Teaching History in Schools, recommendations intended to improve the teaching of history as the core of the social studies. Houghton Mifflin Company responded to the Bradley commission's guidelines with Houghton Mifflin Social Studies 1991, a kindergarten through eighth-grade textbook series. Through this program, Houghton Mifflin met the Bradley commission's guidelines and put its theories into practice. The books have already been adopted by California, Arkansas, Oregon, Indiana, and West Virginia, but it would be premature to say that it has nationwide acceptance. If there were to be a national assessment on social studies, would it reflect the Bradley commission's guidelines or traditional social studies instruction? Would two assessments be needed, so that systems that have adopted the guidelines will not be disadvantaged? Would assessment be delayed until there was a national consensus on the Bradley guidelines? Should there be a national consensus on the adoption of the

Bradley guidelines? These are very practical considerations that must be part of the decision-making process toward a national examination system.

As we know, similar situations exist in mathematics and science as the result of national standards being developed by the National Council Teachers of Mathematics, the Mathematics Science Education Board, and the American Association for the Advancement of Science, as pointed out by Mr. Shanker.

Our authors and editors do go through a consensus process in developing our tests. We consider:

—current emphases in instructional materials, such as textbooks,
—recommendations from national curriculum committees and teacher training specialists,
—critical evaluations and suggestions of classroom teachers and school administrators who use the tests,
—social utility studies in relevant curricular areas,
—comprehensive item tryouts and research studies to determine frequency of error, particularly in language and mathematics,
—independent reviews by professionals from a variety of cultural groups to assess the fairness and appropriateness of items relative to demographics, race, and sex; and
—topic importance, biased on authoritative judgment, instructional trends, and public opinion.

Other test publishers go through comparable processes. The diversity and dynamic process of publishing instructional materials also mitigate against homogeneity and minimize the chances of a close alignment of texts and nationally-available test. Factors against such alignment include that fact that there are many more textbook publishers than there are test publishers, texts and tests are being constantly revised in differing cycles, and tests cover multiple subject areas for which there are likely going to be multiple texts from different publishers. Tests, however, may be selected to more closely correlate to the local curriculum, which may facilitate a school system's use of the test to evaluate progress on local as well as national objectives.

Finally, let us recognize that over the past 20 years there have been real gains in basic reading and numeracy skills, with the greatest gains being realized by minority students. Minimum competency is now nearly universal among high school graduates. That does not mean complacency about higher order skills or the condition of drop-outs, but there is cause to accept that there have been successes. These gains have been recorded on numerous tests, not just on standardized, multiple-choice tests. If minimum national standards are established, and if those standards position our students to meet the challenges of an international economy, perhaps in a few years our assessments will show comparable gains to those shown on basic skills.

To assist the committee's deliberations, I am submitting for inclusion with my statement materials which The Riverside Publishing Company has developed regarding performance and multiple-choice testing formats. Although I can speak only for Houghton Mifflin Company and The Riverside Publishing Company, I am confident that I speak for the commercial textbook and test publishers in general when I say that the industry stands ready to share our publishing expertise with the committee as you strive for a solution to the critical issues we are all addressing.

[Additional copy supplied by Mr. Melody follows:]

CODE OF FAIR TESTING PRACTICES IN EDUCATION

Prepared by the Joint Committee on Testing Practices

The Code of Fair Testing Practices in Education states the major obligations to test takers of professionals who develop or use educational tests. The Code is meant to apply broadly to the use of tests in education (admissions, educational assessment, educational diagnosis, and student placement). The Code is not designed to cover employment testing, licensure or certification testing, or other types of testing. Although the Code has relevance to many types of educational tests, it is directed primarily at professionally developed tests such as those sold by commercial test publishers or used in formally administered testing programs. The Code is not intended to cover tests made by individual teachers for use in their own classrooms.

The Code addresses the roles of test developers and test users separately. Test users are people who select tests, commission test development services, or make decisions on the basis of test scores. Test developers are people who actually construct tests as well as those who set policies for particular testing programs. The roles may, of course, overlap as when a state education agency commissions test development services, sets policies that control the test development process, and makes decisions on the basis of the test scores.

The Code has been developed by the Joint Committee on Testing Practices, a cooperative effort of several professional organizations that has as its aim the advancement, in the public interest, of the quality of testing practices. The Joint Committee was initiated by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education. In addition to these three groups, the American Association for Counseling and Development, Association for Measurement and Evaluation in Counseling and Development, and the American Speech-Language-Hearing Association are now also sponsors of the Joint Committee.

This is not copyrighted material. Reproduction and dissemination are encouraged. Please cite this document as follows:

Code of Fair Testing Practices in Education. (1988) Washington, D.C. Joint Committee on Testing Practices. (Mailing Address: Joint Committee on Testing Practices, American Psychological Association, 1200 17th Street, NW, Washington, D.C. 20036.)

Code of Fair Testing Practices in Education

The Code presents standards for educational test developers and users in four areas:

- **A.** Developing, Selecting Tests
- **B.** Interpreting Scores
- **C.** Striving for Fairness
- **D.** Informing Test Takers

Organizations, institutions, and individual professionals who endorse the Code commit themselves to safeguarding the rights of test takers by following the principles listed. The Code is intended to be consistent with the relevant parts of the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1985). However, the Code differs from the Standards in both audience and purpose. The Code is meant to be understood by the general public; it is limited to educational tests; and the primary focus is on those issues that affect the proper use of tests. The Code is not meant to add new principles over and above those in the Standards or to change the meaning of the Standards. The goal is rather to represent the spirit of a selected portion of the Standards in a way that is meaningful to test takers and/or their parents or guardians. It is the hope of the Joint Committee that the Code will also be judged to be consistent with existing codes of conduct and standards of other professional groups who use educational tests.

A Developing, Selecting Appropriate Tests*

Test developers should provide the information that test users need to select appropriate tests.	**Test users should select tests that meet the purpose for which they are to be used and that are appropriate for the intended test-taking populations.**
Test Developers Should:	**Test Users Should:**
1. Define what each test measures and what the test should be used for. Describe the population(s) for which the test is appropriate.	1. First define the purpose for testing and the population to be tested. Then, select a test for that purpose and that population based on a thorough review of the available information.
2. Accurately represent the characteristics, usefulness, and limitations of tests for their intended purposes.	2. Investigate potentially useful sources of information, in addition to test scores, to corroborate the information provided by tests.
3. Explain relevant measurement concepts as necessary for clarity at the level of detail that is appropriate for the intended audience(s).	3. Read the materials provided by test developers and avoid using tests for which unclear or incomplete information is provided.
4. Describe the process of test development. Explain how the content and skills to be tested were selected.	4. Become familiar with how and when the test was developed and tried out.
5. Provide evidence that the test meets its intended purpose(s).	5. Read independent evaluations of a test and of possible alternative measures. Look for evidence required to support the claims of test developers.
6. Provide either representative samples or complete copies of test questions, directions, answer sheets, manuals, and score reports to qualified users.	6. Examine specimen sets, disclosed tests or samples of questions, directions, answer sheets, manuals, and score reports before selecting a test.
7. Indicate the nature of the evidence obtained concerning the appropriateness of each test for groups of different racial, ethnic, or linguistic backgrounds who are likely to be tested.	7. Ascertain whether the test content and norms group(s) or comparison group(s) are appropriate for the intended test takers.
8. Identify and publish any specialized skills needed to administer each test and to interpret scores correctly.	8. Select and use only those tests for which the skills needed to administer the test and interpret scores correctly are available.

*Many of the statements in the Code refer to the selection of existing tests. However, in customized testing programs test developers are engaged to construct new tests. In those situations, the test development process should be designed to help ensure that the completed tests will be in compliance with the Code.

BEST COPY AVAILABLE

B Interpreting Scores

Test developers should help users interpret scores correctly.

Test Developers Should:

9. Provide timely and easily understood score reports that describe test performance clearly and accurately. Also explain the meaning and limitations of reported scores.

10. Describe the population(s) represented by any norms or comparison group(s), the dates the data were gathered, and the process used to select the samples of test takers.

11. Warn users to avoid specific, reasonably anticipated misuses of test scores.

12. Provide information that will help users follow reasonable procedures for setting passing scores when it is appropriate to use such scores with the test.

13. Provide information that will help users gather evidence to show that the test is meeting its intended purpose(s).

Test users should interpret scores correctly.

Test Users Should:

9. Obtain information about the scale used for reporting scores, the characteristics of any norms or comparison group(s), and the limitations of the scores.

10. Interpret scores taking into account any major differences between the norms or comparison groups and the actual test takers. Also take into account any differences in test administration practices or familiarity with the specific questions in the test.

11. Avoid using tests for purposes not specifically recommended by the test developer unless evidence is obtained to support the intended use.

12. Explain how any passing scores were set and gather evidence to support the appropriateness of the scores.

13. Obtain evidence to help show that the test is meeting its intended purpose(s).

C Striving for Fairness

Test developers should strive to make tests that are as fair as possible for test takers of different races, gender, ethnic backgrounds, or handicapping conditions.

Test Developers Should:

14. Review and revise test questions and related materials to avoid potentially insensitive content or language.

15. Investigate the performance of test takers of different races, gender, and ethnic backgrounds when samples of sufficient size are available. Enact procedures that help to ensure that differences in performance are related primarily to the skills under assessment rather than to irrelevant factors.

16. When feasible, make appropriately modified forms of tests or administration procedures available for test takers with handicapping conditions. Warn test users of potential problems in using standard norms with modified tests or administration procedures that result in non-comparable scores.

Test users should select tests that have been developed in ways that attempt to make them as fair as possible for test takers of different races, gender, ethnic backgrounds, or handicapping conditions.

Test Users Should:

14. Evaluate the procedures used by test developers to avoid potentially insensitive content or language.

15. Review the performance of test takers of different races, gender, and ethnic backgrounds when samples of sufficient size are available. Evaluate the extent to which performance differences may have been caused by inappropriate characteristics of the test.

16. When necessary and feasible, use appropriately modified forms of tests or administration procedures for test takers with handicapping conditions. Interpret standard norms with care in the light of the modifications that were made.

D Informing Test Takers

Under some circumstances, test developers have direct communication with test takers. Under other circumstances, test users communicate directly with test takers. Whichever group communicates directly with test takers should provide the information described below.

Test Developers or Test Users Should:

17. When a test is optional, provide test takers or their parents/guardians with information to help them judge whether the test should be taken, or if an available alternative to the test should be used.
18. Provide test takers the information they need to be familiar with the coverage of the test, the types of question formats, the directions, and appropriate test-taking strategies. Strive to make such information equally available to all test takers.

Under some circumstances, test developers have direct control of tests and test scores. Under other circumstances, test users have such control. Whichever group has direct control of tests and test scores should take the steps described below.

Test Developers or Test Users Should:

19. Provide test takers or their parents/guardians with information about rights test takers may have to obtain copies of tests and completed answer sheets, retake tests, have tests rescored, or cancel scores.
20. Tell test takers or their parents/guardians how long scores will be kept on file and indicate to whom and under what circumstances test scores will or will not be released.
21. Describe the procedures that test takers or their parents/guardians may use to register complaints and have problems resolved.

Note: The membership of the Working Group that developed the Code of Fair Testing Practices in Education and of the Joint Committee on Testing Practices that guided the Working Group was as follows:

Theodore P. Bartell
John R. Bergan
Esther E. Diamond
Richard P. Duran
Lorraine D. Eyde
Raymond D. Fowler
John J. Fremer
(Co-chair, JCTP and Chair, Code Working Group)
Edmund W. Gordon
Jo-Ida C. Hansen
James B. Lingwall
George F. Madaus
(Co-chair, JCTP)
Kevin L. Moreland
Jo-Ellen V. Perez
Robert J. Solomon
John T. Stewart
Carol Kehr Tittle
(Co-chair, JCTP)
Nicholas A. Vacc
Michael J. Zieky
Debra Boltas and Wayne Camara of the American Psychological Association served as staff liaisons

Additional copies of the Code may be obtained from the National Council on Measurement in Education, 1230 Seventeenth Street, NW, Washington, D.C. 20036. Single copies are free.

Bradley Commission Recommendations

1. That the knowledge and habits of mind to be gained from the study of history are indispensable to the education of citizens in a democracy. The study of history should, therefore, be required of all students.

2. That such study must reach well beyond the acquisition of useful information. To develop judgment and perspective, historical study must often focus upon broad, significant themes and questions, rather than short-lived memorization of facts without context. In doing so, historical study should provide content for facts and training in critical judgment based upon evidence, including original sources, and should cultivate the perspective arising from a chronological view of the past down to the present day. Therefore it follows

5. That the curricular time essential to develop the genuine understanding and engagement necessary to exercising judgment must be considerably greater than that presently common in American school programs in history.

4. That the kindergarten through grade six social studies curriculum be history-centered.

5. That this Commission recommends to the states and to local school districts the implementation of a social studies curriculum requiring no fewer than four years of history among the six years spanning grades 7 through 12.

6. That every student should have an understanding of the world that encompasses the historical experiences of peoples of Africa, the Americas, Asia, and Europe.

7. That history can best be understood when the roles of all constituent parts of society are included; therefore the history of women, racial and ethnic minorities, and men and women of all classes and conditions should be integrated into historical instruction.

8. That the completion of a substantial program in history (preferably a major, minimally a minor) at the college or university level be required for the certification of teachers of social studies in the middle and high schools.

9. That college and university departments of history review the structure and content of major programs for their suitability to the needs of prospective teachers, with special attention to the quality and liveliness of those survey courses whose counterparts are most often taught in the schools: world history, Western civilization, and American history.

Houghton Mifflin gratefully acknowledges permission from The Bradley Commission on History in Schools to use material from, **Building a History Curriculum: Guidelines for Teaching History in Schools**. The complete report is available from the National Council for History Education, 26915 Westwood Road, Suite A-2, Westlake, Ohio 44145.

Copyright © 1991 by Houghton Mifflin Company. All rights reserved.

Responding to the Bradley Commission Report on Teaching History in Schools

HOUGHTON MIFFLIN SOCIAL STUDIES

In 1988, the Bradley Commission on History in Schools released *Building a History Curriculum: Guidelines for Teaching History in Schools*, recommendations intended to improve the teaching of history as the core of the social studies. The facing page lists the nine recommendations the Bradley Commission addressed to all citizens who bear responsibility for designing and implementing courses of study in our schools.

Houghton Mifflin Company responded to the Bradley Commission's guidelines with *Houghton Mifflin Social Studies* ©1991, a kindergarten through eighth-grade textbook series. Through this program, Houghton Mifflin endorses the Bradley Commission's guidelines and puts its theories into practice. The following is an analysis on how *Houghton Mifflin Social Studies* responds to those recommendations pertinent to textbooks.

1. That the knowledge and habits of mind to be gained from the study of history are indispensable to the education of citizens in a democracy. The study of history should, therefore, be required of all students.

Houghton Mifflin is committed to providing all students, not just college-bound high school students, with the special skills and knowledge needed to make the decisions democracy will thrust upon them. To that end, a key goal of *Houghton Mifflin Social Studies* is to help all students become responsive and reflective citizens in the twenty-first century.

Houghton Mifflin Social Studies teaches the American ideals and basic principles of democracy that unite all citizens of the United States.

> **While teaching the ideals and basic principles that unite all Americans, *Houghton Mifflin Social Studies* responds to the needs of students of all abilities and backgrounds.**

Throughout the program, students learn about the symbols that shape and reflect our national identity. For example, in grade two, students read about the symbolism of the American flag and the values embodied in the Statue of Liberty. In grade three, students explore the creation of such American symbols as the bald eagle, the Liberty Bell, and "Old Glory."

Houghton Mifflin Social Studies also examines the patriotic values expressed through a variety of historical documents, speeches, and songs. In first grade, for instance, young learners study the Pledge of Allegiance and the song "America." In subsequent grades students consider the values and beliefs expressed in such popular American songs as "America the Beautiful" and "This Land Is Your Land." And key American documents such as the Declaration of Independence and the United States Constitution—to name just two—are presented in their entirety.

In the earliest grades students also examine the beliefs and values that our national and religious holidays reaffirm annually. In kindergarten, students create calendars that identify such holidays as Christmas and Passover. In second grade students learn about Independence Day and Presidents' Day. This is followed up in third grade as students explore the roots of Veterans' Day, Memorial Day, and Arbor Day.

While learning of those beliefs, values, and traditions that shape our national identity, young learners also study the rights and responsibilities of every United States citizen. By second grade, students learn the duties and methods of selecting our political leaders. For example, after studying a lesson on the role and selection of the President, students participate in a decision-making exercise in which they analyze the merits of two plans and vote for the better one.

As the sophistication of instruction increases with each grade, students grapple with more complex democratic procedures. For instance, while studying the turbulent times of Reconstruction, eighth-grade students examine the laborious process of amending the Constitution.

While teaching the ideals and basic principles that unite all Americans, *Houghton Mifflin Social Studies* responds to the needs of students of all abilities and backgrounds. Throughout the program, lessons and chapters inform, motivate, and involve students at all stages of cognitive development and academic ability. Teaching suggestions and strategies contain multiple levels of instructional support that range from teacher-centered activities, for those students who require additional teacher modeling and guidance, to student-centered activities, for those students who can apply, transfer, and extend knowledge and skills independently.

Houghton Mifflin Social Studies is also sensitive to the marked increase in the number of students whose first language is not English. In its commitment to reach all students, this program is designed to facilitate maximum access to social studies concepts for students from a wide variety of cultural and linguistic backgrounds. A unique visual learning strand, prominent instructional visuals, and specific access strategies provide channels of access for students who are acquiring English language proficiency.

The authors and editors of *Houghton Mifflin Social Studies* believe that as citizens of the United States all students need the knowledge and skills gained from the study of history. Therefore, *Houghton Mifflin Social Studies* is designed to instill in *all* students a working knowledge of our democratic process and an appreciation of the values and beliefs that unite all Americans.

2. That such study must reach well beyond the acquisition of useful information. To develop judgment and perspective, historical study must often focus upon broad, significant themes and questions, rather than short-lived memorization of facts without context. In doing so, historical study should provide context for facts and training in critical judgment based upon evidence, including original sources, and should cultivate the perspective arising from a chronological view of the past down to the present day.

The authors and editors of *Houghton Mifflin Social Studies* concur that in order to achieve genuine comprehension of important issues, students need time to explore topics, to practice skills in meaningful contexts, and to relate their learning to the world in which they live. Therefore, *Houghton Mifflin Social Studies* is designed to afford the time required to achieve an in-depth historical focus. Consider, for instance, the portrayal of slave life in

nineteenth-century America. All too frequently, textbooks allot only one or two pages to this topic. Within those pages there is usually only one illustration, which often is a generic handbill advertising a slave auction. The existence of slavery is presented from the perspective of the southern farmer and viewed as a shortcoming of an economic system. While the harshness of slave life may be referred to, it is rarely dealt with in detail. Likewise, the unique African American culture that developed during this time is often overlooked.

Throughout the program, Houghton Mifflin *Social Studies* extends the treatment of vital themes and fully develops the lessons of the past, making the study of history a practical human story.

Houghton Mifflin Social Studies, on the other hand, dedicates entire lessons to the issue of slavery. In *America Will Be* (fifth grade), for example, one six-page lesson that focuses solely on slave life describes slavery from the perspective of Africans. Within this lesson, there are seven visuals, three of which are startling photographs that portray the cruel reality of slavery. Slavery is described as an absolute denial of freedom and human rights.

Striving to develop critical judgment in students, *Houghton Mifflin Social Studies* provides detailed coverage and additional perspective on topics that traditionally receive only cursory review. For example, a typical textbook might state in one sentence that slave families were often broken up with the sale of a father, mother, or child. In *America Will Be*, an excerpt from an autobiography of a former slave vividly describes the grief and sorrow experienced as a family is broken up at a slave auction. While some textbooks may mention that slaves created their own communities and culture, *America Will Be* integrates numerous examples of how slaves adapted foods, crafts, music, and religion to express their African heritage and preserve their own culture.

By providing such detailed coverage and additional perspective, *Houghton Mifflin Social Studies* helps students achieve a level of understanding that promotes genuine comprehension of important issues. As illustrated here, when students take an in-depth look at slavery they gain a deeper appreciation for human rights and better understand the ideals of freedom that have shaped the United States. As students take the time to explore the evolution of African American culture, they develop an appreciation for the multicultural, pluralistic nature of American society.

Whether the topic is slavery in nineteenth-century America or citizenship in the Golden Age of Greece, *Houghton Mifflin Social Studies* integrates knowledge and skills from a wide range of sources and enables students to go beyond the classroom to seek an understanding of the social world. Throughout the program, *Houghton Mifflin Social Studies* extends the treatment of vital themes and fully develops the lessons of the past, making the study of history a practical human story.

4. That the kindergarten through grade six social studies curriculum be history-centered.

In *Houghton Mifflin Social Studies* the following key instructional goals from the scope and sequence establish the program's commitment to a comprehensive history curriculum:

- Create a sense of empathy for the past.
- Analyze the sometimes complex cause-and-effect relationships of ideas and events, recognizing also the effects of the accidental and irrational on history.
- Recognize the interrelatedness of geography, economics, culture, belief systems, and political systems within history.

In the primary grades, students quickly move beyond their own world to broaden their sense of time and place. In kindergarten, young learners are introduced to the concept of change over time through the comparison of posters depicting, for example, a neighborhood long ago and today. The program also gives students new insight into their communities and helps them make connections to the larger world. In first grade, for example, students examine the growth of a town into a city and observe the role of a bridge and transportation in linking the city to the wider world.

Throughout *Houghton Mifflin Social Studies* a variety of lesson formats invite students to explore the geographic, social, and economic aspects of the world. At every grade, literature, biographies, and primary sources introduce young learners to people and events in other times and places.

By the third grade, students begin their study of the United States as they learn about its early people and their cultures. In the fourth grade, students continue their exploration of the nation as they survey each of its five geographic regions from an economic perspective, both in early times and in the present day.

In the upper grades, students continue to study geography, economics, anthropology, and political science as well as the expressive arts and the humanities within the context of history. In the fifth grade, students examine United States history from the appearance of the first people up to the Civil War. In the eighth grade, students review early America and focus in depth on the period from the Constitutional Convention to the First World War.

Houghton Mifflin Social Studies offers two years of world history in grades six and seven. In grade six students study the ancient world from its beginnings through classical civilization up to the decline of the Roman Empire. In the seventh grade students expand on their knowledge by studying the civilizations of Africa, China, Japan, the Americas, and Europe from the fall of Rome through the Middle Ages to the Enlightenment.

Throughout the program, history provides an organizing framework to explore the many disciplines that make up a social studies education.

6. That every student should have an understanding of the world that encompasses the historical experiences of peoples of Africa, the Americas, Asia, and Europe.

Houghton Mifflin Social Studies responds to the fact that students today are living in a nation that is becoming more culturally diverse and in a world that is increasingly interdependent. Its commitment to a well-rounded world view is established in the following key instructional goals from the scope and sequence:

- Understand world regions and their historical, cultural, economic, and political characteristics.
- Understand the complex relationship and interdependence that exists among the world's nations.
- Become familiar with the basic ideas of major religions and ethical traditions of other times and places.

Houghton Mifflin Social Studies examines the topics of pluralism, interdependence, and the multicultural nature of our society at every grade level. For example, *Some People I Know* (second grade) presents a complete unit on the multicultural nature of the United States through a study of real families from five different cultural heritages: Cambodian, Mexican American, German, African American, and Navajo.

America Will Be (fifth grade) initiates a study of United States history by looking at the theme of cultural diversity. Focusing on New Orleans as a microcosm of the United States, the first chapter describes the immigration process—both voluntary and forced—that brought so many different groups to this nation. Using primary source materials, engaging biographies, and literature to explore a variety of immigrant experiences, the chapter examines the benefits and challenges of living in a land of many cultures. With this preamble, students embark on a study of our nation's history from a pluralistic perspective. *America Will Be* and *A More Perfect Union* (eighth grade) devote significant attention to such issues as immigration, prejudice, and civil rights in United States history.

> **Through *Houghton Mifflin Social Studies*, students gain an understanding of the unique nature of world cultures and the characteristics shared by all civilizations.**

Houghton Mifflin Social Studies also takes an honest and balanced look at cultural diffusion and cultural conflict in other times and places. *A Message of Ancient Days* (sixth grade) and *Across the Centuries* (seventh grade) dedicate their entire text to teaching the cultural heritage and contributions of the peoples of Africa, Asia, the Middle East, the Americas, and Europe. In grade six, while pursuing an in depth study of the ancient civilizations of Mesopotamia, the Nile Valley, the Indus Valley, and the river valleys of China, students gain an understanding of the unique nature of world cultures and the characteristics shared by all civilizations. In grade seven, while studying the accomplishments of the Muslims, the Aztecs, the Bantu, and the people of the Kongo—to name a few—students learn how the exchange of ideas, scientific developments, and artistic achievements help cultures grow and develop.

7. That history can best be understood when the roles of all constituent parts of society are included; therefore the history of women, racial and ethnic minorities, and men and women of all classes and conditions should be integrated into historical instruction.

Houghton Mifflin believes that a truly democratic education demands that we acknowledge not only the familiar heroes and noted politicians but also the common people and unheralded leaders who worked in many roles to shape history. From its inception, *Houghton Mifflin Social Studies* was built around the principles of a gender-sensitive, multicultural education. These principles are set forth in the following key instructional goals from the program's scope and sequence:

- Develop an awareness of the structure of social classes and the changes in status of women and racial and ethnic minorities in U.S. society and other societies.
- Develop an appreciation for the multicultural, pluralistic nature of U.S. society.
- Comprehend the history of women, minorities, and the full range of social classes, not just the history of the elite or the notable individual.

To achieve these goals, *Houghton Mifflin Social Studies* integrates the content with numerous perspectives.

Throughout the program, the authors weave into the narrative of history the voices and experiences of individuals of many ethnic groups and from all stations of life. In the early grades, these include biographies of such individuals as Yoshiko Uchida, a prominent author of children's literature; Roberto Clemente, the humanitarian and hall-of-fame baseball player; Alexa Canady, the noted surgeon; and Eleanor Roosevelt, the popular First Lady and renowned diplomat. In later grades, students read the experiences of Luther Standing Bear, the Sioux Indian Chief and author; Jackie Cochran, the world class pilot; Samantha Smith, the international peace activist; and Jessie de la Cruz, the labor organizer and politician—to name a few.

Because the authors and editors of *Houghton Mifflin Social Studies* believe students should hear the voices of the people who made and lived history, the program also features a profusion of primary sources and literature. Students read from the slave

BEST COPY AVAILABLE

narratives of Olaudah Equiano, Solomon Northup, and Frederick Douglass, the speeches of Black Hawk, Chief Joseph, Martin Luther King, Jr., and Pablo de la Guerra, the diaries and journals of Ellen Kiskiyama, Charlotte Forten, and an anonymous Vietnamese refugee.

> **Throughout the program, the authors weave into the narrative of history the voices and experiences of individuals of many ethnic groups and from all stations of life.**

In addition, students gain diverse perspectives on history through such literature as "Carrying the Running-Aways," a true slave narrative about the Underground Railroad; "How the Turtle Beat the Rabbit," a Cherokee folktale; "A Jar of Dreams," Yoshiko Uchida's story about the daughter of Japanese immigrants; and "The Voyage of Sinbad the Sailor," an Arabian tale from the collection *One Thousand and One Arabian Nights*. In fact, the inclusion of multicultural literature begins even in kindergarten with two South African stories: *Not So Fast, Songololo* and *Jafta—The Journey*.

Throughout the program, *Houghton Mifflin Social Studies* instills in students an empathy for people of other times and places. For example, the personal insights provided by "A Moment in Time" characters help students identify with common people as they journey through history. In the earliest grades, students observe "A Surgeon" as she prepares for an operation, a "Pioneer on the Oregon Trail" as she collects supplies on the banks of the Platte River, and "A Vaquera" at a roundup near San Antonio. In later grades, students examine "A School Reformer" at a rally in Virginia, "A Roman Engineer" as he works on a road outside the city of Pisa, and "A Kongo King" in the capital of Mbanza.

In every grade, students take an in-depth look at people like themselves, whose decisions and actions provided the crucial grassroots support that fuels important historical movements. By including the contributions of everyday men and women, immigrants, and various ethnic groups, *Houghton Mifflin Social Studies* provides vital insight into the strength of our democratic society.

Criteria for All History Courses and Curricular Patterns

To achieve the desired history-enriched curriculum, the Bradley Commission endorses alternate course sequences at the elementary, middle, and high-school levels. All of the Commission's recommended curricular patterns, while different, evolve from a shared educational philosophy. Throughout *Building a History Curriculum*, the Commission stresses specific criteria that should be maintained in every history course and curricular pattern. The following is an analysis of how *Houghton Mifflin Social Studies* integrates three of these criteria throughout its program.

An Interdisciplinary Subject

One criteria that is fundamental to all of the Commission's recommended curricular patterns requires that history be taught as an interdisciplinary subject. The report states, "History, by its nature, is an interdisciplinary subject. It should never be reduced to a thin recital of successive dates and facts, but carry what has been called 'thick narrative,' which combines lively storytelling and biography with conceptual analysis drawn from every relevant discipline."

Houghton Mifflin Social Studies is built on the premise that social studies should be taught as just that—a social study—a study of people in their social world. A comprehensive social study has to consider how people live, work, worship, and play together. It has to consider how they govern themselves and how they make important decisions. Attempting to represent all of the elements of life, the program presents knowledge of the social world in a holistic and interrelated manner.

Houghton Mifflin Social Studies prides itself on its richly integrated content that draws from history, geography, economics, anthropology, sociology, psychology, political science, and philosophy as well as the expressive arts and the humanities—literature, music, art, and architecture.

> ***Houghton Mifflin Social Studies* provides students with a multidimensional view of the world and a practical understanding of the ideas and issues that have shaped history.**

A lesson in *This Is My Country* (fourth grade) illustrates the point. In teaching about the gold and silver rushes of the mid- and late-nineteenth century, the lesson begins with an excerpt from Klondy Nelson's *Daughter of the Gold Rush*. The excerpt provides a colorful firsthand account of a mining camp as seen through the eyes of a young girl. This is followed with a concise history of the West's many "rushes" and the geographic factors that shaped the migration to and the settlement of mining camps.

The text then describes the evolution of mining technology, the legal process of filing a claim, and the social makeup of mining camps and towns. The lesson's "A Moment in Time" describes "A Forty-niner" as he pans for gold. This feature describes the miner's tools, clothing, diet, background, and dreams. Additional exercises and strategies connect the gold rush to science by describing the natural formation of placer and mining of it. Additional exercises also challenge students to consider some of the ethical issues faced by miners as they worked in the gold fields. The lesson closes with a story from *By the Great Horn Spoon!*, an exciting historical novel by Sid Fleischman that describes life during a gold rush.

By maintaining an interdisciplinary balance, *Houghton Mifflin Social Studies* provides students with a multidimensional view of the world and a practical understanding of the ideas and issues that have shaped history.

Less Is More

Another criteria that is basic to all of the Commission's curricular patterns is the notion "less is more."

For years the traditional social studies curriculum has endorsed a survey approach to history, a curricular pattern that attempts to teach a general overview of history. For example, the traditional United States history survey charts history from the Age of Exploration to the present in one school year. But, by trying to depict too much, survey courses are unable to present information in a thorough manner. Some periods, especially the twentieth century, receive only definitional review, with many important concepts being completely overlooked.

Sensitive to this shortcoming, the Bradley Commission states, "Has the notion that 'less is more' been considered, as themes, topics, and questions are selected? The amount of time required to achieve student engagement and genuine comprehension of significant issues will necessitate leaving out much that is 'covered' by the usual text."

Sharing this concern, the authors and editors have designed *Houghton Mifflin Social Studies* as a continuum with a narrower chronological focus. Instead of repeating general overviews of history, students take a detailed multi-dimensional look at specific periods of history. For example, instead of studying the entire span of United States history in the fifth grade, then again in the eighth grade, and yet again in the eleventh grade, students using *America Will Be* (fifth grade) focus on United States history from the appearance of the first people up to the Civil War. Then in *A More Perfect Union* (eighth grade), students review early American history and analyze United States history from the Constitutional Convention to events surrounding World War I. These comprehensive studies combine to create a solid foundation for the study of twentieth-century United States history in *History of the United States* (eleventh grade).

> ***Houghton Mifflin Social Studies* gives students the time to make an in-depth, multi-perspective study of history.**

By narrowing the chronological focus, *Houghton Mifflin Social Studies* gives students the time to make an in-depth, multi-perspective study of history. It also gives students the time to achieve genuine comprehension of significant issues, issues that will reappear in future courses and throughout their lives. For instance, while studying Puritan society in *Time Will Be*, students examine in detail the issue of dissent and protest. Through this in-depth study, students achieve a more accurate understanding of seventeenth-century America. In addition, because students have had time to understand the roots of dissent in United States history, they will have greater perspective when they study the enactment of the Sedition Act of 1918 in *A More Perfect Union* and the civil rights protests of the 1960s in *History of the United States*.

Education for Democratic Citizenship

A third criteria that reverberates through all of the Commission's recommended curricular patterns is a commitment to the education of democratic citizenship. The Commission states, "Like the Committee of Ten and all of its successors, the Bradley Commission is concerned with the central role of history for civic education. The social studies have always taken education for democratic citizenship, and for the public good, as their primary aim."

In its response to the Bradley Commission's first of nine recommendations, *Houghton Mifflin Social Studies* establishes its commitment to instilling in students an appreciation of our shared national identity and a working knowledge of the democratic process. But the education of democratic citizenship must go beyond supplying a historical grasp of our common political vision. If a goal of education is to promote active, caring, intelligent citizenship, students of all ages need to be taught civic responsibility in conjunction with sound decision-making exercises.

> **By studying democratic principles and civic responsibility in conjunction with sound decision-making exercises, students develop the solid knowledge and practical skills needed to be active, caring, intelligent citizens.**

In the earliest grades of *Houghton Mifflin Social Studies*, young learners study rights and responsibilities at a very basic and personal level. In *The World I See* (kindergarten), students practice getting along and working together through numerous group tasks and role-playing exercises. A series of activities in one lesson helps students identify rules and understand

why it is important to follow rules both in school and outside of school. Students then make signs that help others follow rules. In a lesson on friendship in *I Know a Place* (first grade), students discuss ways people show friendship and respect for others, and then they apply lesson examples to classroom situations.

Throughout *Houghton Mifflin Social Studies*, civic responsibility is also conveyed through graphic portrayals of virtue, courage, and wisdom. For example, in *Some People I Know* (second grade), students learn about Roberto Clemente. While his achievements as a baseball player are noted, it is his selfless commitment to others and his humanitarian deeds that are highlighted. Likewise, as students study Eleanor Roosevelt, they learn about her impressive career as a diplomat as well as her struggle for human rights.

Against this backdrop of inspirational role models, students are challenged to consider the effects of technological, economic, and cultural change, and the resulting choices we must make. Presenting meaningful knowledge of the social world is a fundamental concern of the authors and editors of *Houghton Mifflin Social Studies*. From the outset we believed that instruction needs to focus on information that shows students how richly all the elements of social studies touch everyday decision making. We believe that social studies should not only allow us to understand how others have made decisions but also inform our own decision making and our awareness of current issues and social concerns.

By studying democratic principles and civic responsibility in conjunction with sound decision-making exercises, students develop the solid knowledge and practical skills needed to be active, caring, intelligent citizens.

In closing, the authors and editors of Houghton Mifflin support the Bradley Commission's efforts to elevate history instruction in our schools. We share the belief that all children in a democracy deserve the knowledge and understanding that history imparts. The formation of a nation of powerful decision makers begins in the classroom, and its ramifications will be felt by future generations. A future in which today's students have solid knowledge, know when and how to use skills, and possess an understanding of involved citizenship will be a future in which the principles of democracy will grow and flourish. Houghton Mifflin is proud to be a part of the process of creating literate citizens, of forming powerful decision makers who care both for our nation and for civilization itself.

■ For further information about *Houghton Mifflin Social Studies* contact the Regional Office for your area.

7055 Amwiler Industrial Dr.
Atlanta, GA 30360
Tel: 404-449-5881
1 (800) 241-9612 (Outside GA)
1 (800) 733-8786 (Within GA)

13400 Midway Rd.
Dallas, TX 75244
Tel: 214-980-1100
1 (800) 733-2828 (Outside TX)
1 (800) 442-5628 (Within TX)

1900 S. Batavia Ave.
Geneva, IL 60134
Tel: 708-232-2550
1 (800) 323-5663 (Outside IL)

925 E. Meadow Dr.
Palo Alto, CA 94303
Tel: 415-857-0211
1 (800) 222-8460 (Outside CA)
1 (800) 982-6111 (Within CA)

101 Campus Dr.
Princeton, NJ 08540
Tel: 609-452-0200
1 (800) 733-7075 (Outside NJ & ME)

THE COLLEGE BOARD REVIEW
FALL 1990

Multiple-Choice and Its Critics

Are the "Alternatives" Any Better?

by Lawrence Feinberg

IN Vermont this fall every student in the fourth and eleventh grades will be issued a special folder. It will contain a "portfolio" of the students' best writing to be graded in the spring not only by their own teachers but also—in some cases—by teams of outside evaluators.

In New York last year all fourth graders took the same unusual test—manipulating lab equipment. Their proficiency was rated according to a common scale.

In Maine all eighth graders had to write an essay about the same photograph—a satellite picture of a forest fire near Los Angeles—telling what it threatened and offering a plan to handle it.

These three examples and dozens more like them around the country are the visible signs of a heated debate—both academic and political—about testing. Multiple-choice tests, which have expanded in number and influence for more than a decade since the late 1970s, have come under vehement attack. In their stead, their detractors propose "alternative assessments"—essays, demonstrations of performance, and collections or "portfolios" of student work.

These alternatives, their proponents say, are more "authentic" than the familiar fill-in-the-bubble exams and more fair; they will shift the curriculum to higher-order thinking skills and encourage better teaching.

However, even though a number of states are experimenting with the new assessments—most notably California, Connecticut, and Vermont—the alternatives themselves are far from fully developed. The issues they

In addition to traditional multiple-choice design, students may soon face an array of new test formats.

raise are important but the effects they may have are far from clear.

Ironically, multiple-choice tests themselves came into widespread use over the past four decades partly in an effort to achieve the very fairness that the critics say they lack. They too were once an "alternative"—to traditional essay exams. The topics of those exams were derided as arbitrary by an earlier generation of test critics; their grading was denounced as inconsistent and influenced by irrelevancies, such as neat handwriting.

Of course, attacks on standardized, multiple-choice testing are nothing new. A major assault was mounted in the 1970s, led by the National Education Association (NEA). In1972 the NEA even called for a moratorium on all standardized testing. Later, the teachers' group was joined by Ralph Nader, the consumer activist, and by the National Association for the Advancement of Colored People (NAACP).

Despite extensive attention, the critics were routed not so much by school officials and testing organizations as by the genuine grass roots push for minimum competency exams as an antidote to "worthless diplomas." Also, in 1983, the National Commission on Excellence in Education used test data as the linchpin of its report that American schools were beset by a "rising tide of mediocrity." That ushered in more testing to set standards for students and to hold schools accountable.

As Barbara Lerner, the psychologist and lawyer, has noted, many of the arguments now used against multiple-choice testing are the same as those used in the earlier assault and many of the same people are making them. However, there are several new elements in the ferment.

SOME of those who earlier had defended testing—most notably Albert Shanker, the president of the American Federation of Teachers—have now turned against multiple-choice tests. And two new arguments have been added to the critics' arsenal: Multiple-choice tests test facts, not thinking, and preparing for them encourages bad teaching. In addition, at least some of the critics—though by no means all—are really looking for standardized alternatives to the format of multiple-choice tests. By contrast, in the 1970s the NEA and Nader extolled the virtues of teacher judgment and grades without any outside exams.

As Lerner noted, the traditional arguments against multiple-choice tests boil down to three points:

1. The ability of tests to predict future performance in class or outside is far from perfect.

2. The tests don't measure many traits that are important, such as persistence and the ability to organize time and work.

3. The tests are biased against minorities, the poor, and, perhaps, women.

The first two points are true but far from telling; the third is false and "has been disproved by solid, empirical data over and over again," Lerner said.

In fact, the first two points are arguments against any type of timed, standardized exam—multiple-choice, essay, or defined performance. There is no evidence that any test format is better than any other in foretelling the future or divining character or motivation.

However, as a National Academy of Sciences committee concluded in 1982, there is a wide body of evidence that multiple-choice tests "can be useful predictors" of job performance and college grades if they are properly developed and interpreted. This is so because the intellectual skills and abilities that the tests measure are, in fact, important factors in determining how well students and workers perform.

In general, the tests predict less well than grades, but standardized tests plus grades predict much better than grades alone because grades vary enormously from teacher to teacher.

Of course, the tests don't measure everything important. Character and creativity certainly count, as Lerner says, and neither can be gauged by a standardized test. However, neither quality really can be substituted for the academic abilities that tests do measure. To succeed in college a student needs preparation for college-level work; no amount of zeal or creative flair can make a student at a fifth-grade level of achievement capable of handling its rigors. And academic tests can clearly show where students stand academically.

But academic ability alone is not sufficient for academic success. Effort and self-discipline are needed too. This, of course, is the fundamental reason why any academic test cannot make perfect predictions.

However, according to the National Academy of Sciences report, the tests do not discriminate by race or sex or social class. Regardless of these factors, students with

Lawrence Feinberg, assistant director for reporting and analysis for the National Assessment Governing Board, covered education for many years for the Washington Post. This article represents his own views and not necessarily those of NAGB.

Teaching to the test, as well as a narrowing and distortion of curriculums, are recent charges from critics of multiple-choice.

similar test scores get similar grades in similar courses. Indeed, the only slight exception to this rule is an unexpected one: Blacks with a given test score tend to get lower grades than whites or Asians with the same test scores. In other words, the multiple-choice tests tend to overpredict slightly for blacks, not to underpredict, as the critics contend.

Overall, women with given test scores get higher grades than men, but if the sexes are matched by the courses taken, the difference begins to disappear. More high-scoring men enroll in science and math, where grades tend to be lower, than in arts and humanities.

As to whether males or females perform better on multiple-choice or essay exams, the evidence is inconclusive, according to a major new study by Ross E. Traub and Katherine MacRury of the Ontario Institute for Studies in Education. Females do score relatively higher on essay exams in Great Britain, Ireland, and Australia, they report, where both essay and multiple-choice questions are used. The same is true on the California Bar exam, which has one-hour and three-hour written sections, as well as a multiple-choice test. But the authors note another body of research that shows females write more neatly than males and that neat handwriting is associated with higher essay scores.

There simply is no evidence that any test format is better than any other in foretelling the future or divining character or motivation.

THERE seems to have been little systematic research comparing score gaps between blacks and whites on essay questions and multiple-choice. A study of the California Bar exam indicated that adding the lengthy written performance test in 1984 did not narrow the gap between whites and minorities. The pass rate for all groups went down when the exam changed, but the proportion of blacks passing remained more than 40 percentage points below that of whites. Indeed, the study by Stephen P. Klein of the RAND Corporation says that if the performance test scores were adjusted to take into account the lower reliability of performance test grading, the racial gap would be even wider than on multiple-choice.

A recent report by the National Assessment of Educational Progress (NAEP) on its 1988 writing exam points to the same conclusion. Students were given a half hour to answer each exam question instead of the 15-minute time limit imposed in the past. The impact of having more time was substantial, raising the average performance of white students while blacks and Hispanics were little changed. Thus, the racial gap appeared to widen, compared to that on the shorter writing tests, where it had been just about the same as on NAEP's mostly multiple-choice exams in other subjects.

The explanation, according to NAEP officials, was that giving more time allowed "pretty good" writers to organize their thoughts better and write better prose, while those who could produce only "minimal" essays didn't write any better. According to the test data, pro-

portionately more whites than blacks were in the "pretty good" category, so more whites than blacks moved up. The same thing happened several years earlier in South Carolina when that state lengthened the time for answering each writing test question.

One of the main criticisms in the current wave of opposition to multiple-choice tests is that the tests themselves are too simple. As the *New York Times* put it, "Standardized multiple-choice tests have drawn increasing fire as too simplistic, measuring the ability to recognize knowledge rather than the ability to think and solve problems."

There are two problems with this accusation. First, a test of knowledge is not necessarily simplistic. For good reasons it is a major part of the sophisticated multiple-choice licensing exams for physicians and engineers, as well as of tests of fifth-grade science. Even though knowledge may be in disrepute among some education theorists (though not among the public), no amount of reasoning about a subject can make much sense unless students first know something about it. Even the critics concede that multiple-choice tests are an excellent way to test what students know, a consideration not to be lightly disregarded.

Perhaps the most serious accusation leveled against multiple-choice tests is that they discourage teaching without writing.

Of course, many multiple-choice tests are simple. Most minimum competence tests for high school graduation are pegged at a junior high level.

Secondly, many multiple-choice tests are more than *just* tests of knowledge. In fact, the most widely known multiple-choice exam, the Scholastic Aptitude Test, tests very little knowledge; it is almost completely a test of analytical and reasoning ability at quite complex and sophisticated levels. The multiple-choice questions on Advanced Placement exams—about half of each test—often require sharp analysis with a command of facts. Many other standardized tests, particularly at the high school level, also probe the ability to draw fair inferences and reach sensible conclusions.

Of course, many multiple-choice tests are simple. Most minimum competency tests for high school graduation are pegged at a junior high level. But the decision to have such easy exams is a political one, reflecting a fear of "too many" failures. An exam's level of difficulty is not a function of its format. Indeed, New Jersey and a few other states have moved to higher standards on harder, more analytical multiple-choice exams.

The nationally standardized tests for elementary and secondary schools vary considerably in difficulty and the amount of reasoning demanded by their questions. Unfortunately, the ones most widely used by public schools are at the low end of the scale for obvious, self-serving reasons. By contrast, most private schools—and a few public school districts—use much harder, more analytical tests sponsored by the Educational Records Bureau.

At the very highest levels of complexity, of course, an essay or written performance question can show higher performance than a question that is multiple-choice. After all, the answers aren't limited by the choices presented; there is no "ceiling" on creativity or insight. But for most testing in most schools, there is considerable room for more challenging material on multiple-choice tests. It is an educational and political decision by school officials whether or not to administer such challenging exams, not an issue of test format.

Conversely, not all free-response questions are complex. Many are simple, fill-in-the-blanks affairs. Even written essays can be rigid and simplistic. For example, Maryland's writing test for high school graduation requires a highly structured essay. The schools are told

what that structure will be, and so they teach it. Some English classes stop regular work two weeks before the exam while teachers teach the essay formula. Most students do well even in ninth grade—the first time they take it—but its impact on student writing outside the "formula essay" seems minimal. Of course, that is precisely the criticism often made of highly structured multiple-choice tests whose exact parameters—and sometimes exact questions—are well-known to teachers and schools.

The charge that multiple-choice tests have become the focus of "teaching to the test" and have distorted and narrowed the curriculum is a serious one. There is some evidence that it is happening, particularly in schools where achievement has been low. Sales of test preparation books have boomed; in some districts even kindergartners are taught to fill in answer sheets.

How widespread and how harmful these practices are is highly uncertain. Over the past decade, performance in the basic skills that schools and tests have stressed does seem to have improved—not only on commercial tests but also on the National Assessment, which is given on a sample basis (though NAEP gains are more modest). Although improving basic skills is not sufficient, it's not to be sneered at either, especially compared to the general deterioration of all school achievement that took place from the mid-1960s to mid-1970s.

Also, the practices that critics often dislike seem to be a result of how multiple-choice tests are normed, administered, and used, and not an inevitable consequence of the format itself. A different type of test may well be subject to the same problems. If a performance test were repeated for as many years as multiple-choice tests often are, students might well be trained to perform the tasks tested and know little else about a field.

Actually, there is considerable irony in the way many school districts prepare their students for the tests because there is very little evidence that general "test-wiseness" has much to do with test results. What may well have an impact is training for the specific topics and question forms themselves or very close facsimiles. And that, as the president of Friends for Education Dr. John Cannell has charged, amounts to something like cheating.

At the very highest levels of complexity, of course, an essay and written performance question can show higher performance than a question that is multiple-choice. After all, the answers aren't limited by the choices presented; there is no "ceiling" on creativity or insight.

The solution is to regularly give new tests, to create new norms based on actual performance, and to impose tighter test security. All that may cost more money, but it can be done with multiple-choice exams.

In large part, these problems of testing have been created by the way standardized tests are used to evaluate schools and school systems, not just individual students. There is no way, however, that the genie of publicity and accountability can be put back in the bottle. Nor should it be. Parents and the public really ought to know how their schools compare to others. However, this can be done accurately through sample testing from a large bank of questions (mostly multiple-choice), as in the National Assessment. These assessment exams can randomly test broad swatches of curriculum, making it difficult to cheat and encouraging broad learning, rather than the narrow preparation encouraged by narrow tests—whether multiple-choice or some other type.

Indeed, multiple-choice questions have the great advantage of being able to test a much broader range of topics, compared to essay or performance examinations.

According to the critics, multiple-choice tests have harmed the curriculum in another way. By testing without writing they have encouraged schools to *teach* without writing, and this has been a serious loss. Actually, the best evidence is that schools where students write often get higher scores on multiple-choice tests—certainly on the SAT-verbal and on other tests too that require analytical ability. Tests that include writing undoubtedly will encourage more of it, and writing itself, in contrast to every other subject, can only be tested fully by written exams.

But this too is not without pitfalls. Even when great pains and expense are taken to achieve uniformity, any grading that involves human graders and judgment is more prone to inconsistency than machine scoring. And because fewer topics can be tested in a particular length of time, any written exam necessarily samples a smaller slice of what students know, unless an exam gets to be unreasonably long. Compared to multiple-choice tests of similar length, written exams more arbitrarily emphasize one topic or another with which a student may (or may not) be familiar.

(continued on page 30)

Multiple-Choice

(continued from page 17)

THERE may also be some confusion over what "authentic" tests actually test. For example, on a reading test that requires written responses rather than multiple-choice, it may be unclear whether reading comprehension or writing ability is being measured. In an open-ended science or math test that stresses strategies and problem-solving, the test may be measuring general ability—the old IQ of intelligence testing—rather than mastery of content.

Teacher observations and checklists have similar problems. When there is no written product to be evaluated, the scoring of these tests is much less consistent than the grading of essay exams.

Ruth Mitchell, associate director of the Council for Basic Education, noted this problem firsthand in a recent observation of performance assessments being developed in Great Britain. Although she admired the assessment tasks themselves, Mitchell wrote that "reliability was compromised because during the course of the day, the teacher became more proficient at introducing the task (dropping paper helicopters with different wing lengths to see which fell fastest) to successive groups, and the children benefited from watching the earlier groups."

With these problems and others the British have decided to delay introduction of the new tests by a year. A similar delay for many of the same practical reasons has occurred in California, even though that state's Education Summit resolved last December that "all multiple-choice tests should be eliminated."

"Maybe that was hasty. Maybe it's ill-conceived. Maybe it won't happen anyway," Dale Carlson, director of the California Assessment Program, said last spring. "But we're trying."

The portfolio evaluations face another obvious problem: Whose work is being evaluated? For teachers evaluating their students' writing, the uncertainty probably is not a concern, though even in this case (as in science fair projects and college admission essays) parents may help, and students with well-educated parents—mostly in the middle class—will probably be helped the most. However, if portfolios are used to evaluate schools or school systems there can be no assurance that teachers haven't improved the work, thus confounding the assumption that it represents performance.

If multiple-choice tests are abolished and replaced by some other type of uniform, reliable exams, who does well or poorly is not likely to be much different from what it is now. On the California Bar exam, the largest program so far to have incorporated performance testing, the rank order of applicants is nearly the same on the performance, essay, and multiple-choice sections. Low-scoring students score low on all three parts; high-scoring candidates score high on all three.

According to a new study, the same thing is true on the free-response and multiple-choice parts of the Advanced Placement computer science exam. Several similar studies on other tests have yielded similar conclusions, although low-performing students seem to do relatively worse on essay exams because the performance ceiling is raised (although this may be counterbalanced by lower reliability in scoring).

Thus, for two of the main purposes of testing—placement (or admission) and evaluating institutions—the form of the test seems to make little difference.

For the third main purpose—the direct improvement of learning through individual diagnosis and instructional feedback—the essay and other free-response exams certainly can tell more than multiple-choice. They can show *how* students get an answer or make a mistake, not just whether the answer is right or wrong.

This may well be of more help to teachers than multiple-choice test scores. But teachers themselves already give many such tests. Whether producing and scoring them according to a state or national standard makes much difference remains to be seen.

Although the denunciations are stern and the controversy sharp, there may soon be a place in large-scale school testing for both multiple-choice and some of the performance alternatives (though few of the alternatives are actually ready for use). But despite fervent claims, it seems unlikely that any change in test format will have as much impact as harder test content and higher test standards might have.

What's tested and what's expected matters much more than which way the testing is done. ●

EDUCATIONAL PERFORMANCE TESTING:

OR, DO WE REALLY NEED ANOTHER KIND OF TEST?

by F. L. Finch

California Educational Research Association
Santa Barbara, CA November 15, 1990

The following ideas about educational performance testing resulted from practical experience in developing, administering, and scoring the research editions of the Arizona statewide performance testing program. Other issues emerged during the development of Riverside's Integrated Literature and Language Arts Portfolio Program. One basic principle applied to these projects was that the use of multiple-choice item formats would be minimized. At the beginning of each project, we assumed that the interest in alternative forms of assessment reflected a genuine desire for qualitatively different tests.

The educational community has expressed a strong interest in alternatives to, or supplements for, existing tests. This interest comes from, at a least in part, a belief that the format of traditional multiple-choice tests places limitations upon what can be measured. This is true. It is also true that the seriousness of these limitations has been grossly overstated by some.

It is also true that the advantages of various alternatives to multiple-choice tests have sometimes been exaggerated. This sentiment was recently expressed by Rothman (1990) who stated, "Some skeptics are charging, however, that it is premature to throw out traditional tests in favor of the new method... Advocates of performance assessments may have inflated their value by claiming that they would improve teaching and spur instruction on higher-order skills." (p. 1) Even so, most test publishers believe that we have the capability of producing a new generation of educational assessments, and we are in the process of providing them.

The phrase "LAW OF THE INSTRUMENT" is common usage among experimental psychologists. In brief, the law of the instrument pertains to the common error of attempting to use a favorite measure for all possible purposes. The term also includes the tendency of researchers to find excuses to use specific instruments. The traditional elaboration on the law of the instrument is: "If you give a child a hammer, he or she will find a lot of things that need pounding."

42-921 0 - 91 - 6

The law of the instrument reminds us that machine-scored, multiple-choice, tests should not be viewed as the solution to every measurement question. They are efficient and accurate, but they are not a universal panacea. The point that escapes many critics is that no reputable test publisher recommends objective tests for some of the purposes that the critics often cite as examples of inappropriate uses of tests. For example, those who assert that students cannot be described by a single score ignore the fact that no knowledgeable developer or user would do this. (Standard 8.12)

In the 1970's, criterion-referenced tests (Berk 1984) and life-skills tests (Finch et. al. 1980) were viewed as alternatives to traditional educational tests, but these tests typically retained the multiple-choice item format. Many CRT and life-skills concepts were later integrated into mainstream achievement tests.

In the last year or so, the educational community and some political action groups have become interested in "performance testing." Many articles have been written about "authentic assessments" as "alternatives" to "standardized" tests. This single sentence illustrates three problems which pertain mostly to semantics. Semantics, as we all know, deals with the relationships between symbols and what they represent. The discussion of testing issues has been clouded by the use of vague or inappropriate terms.

First, criticisms of "standardized" tests (see Beyond Standardized Testing, 1988) pertain, it is assumed, to multiple-choice tests that are scored by computers. Anastasi points out that "standardization implies uniformity of procedure in administering and scoring the test. If the scores obtained by different persons are to be comparable, testing conditions must obviously be the same for all." (1988, p. 25) Informal assessments can be useful, but it is assumed that any assessment that is designed for widespread use must be based on uniform procedures. That is, it must be a standardized test. (also see Finch, 1990.)

A second problem pertains to the use of the term "alternatives." Alternatives to objective tests have always been available. Test publishers and educators recommend that this information be used to supplement objective test data. The semantic problem is that these alternatives are recommended as replacements for existing tests. Alternative are always welcome, but the notion that these "new and improved" tests are better than conventional measures requires more evidence than the assertions and assumptions that have been provided to date. The phrase "tossing the baby out with the bath water" comes to mind when the need for alternatives (i.e. replacements) is suggested.

The third semantic problem is that the various labels associated with the new tests appear to be more closely related to advertising than measurement. For example, the term "authentic assessment" (Archibald and Newmann, 1988) sounds great but is not very descriptive. The term "portfolio assessment" has also been suggested and, although it is a bit closer to the mark, seems more appropriate for describing the receptacle than its contents. Figure 1 below illustrates this point.

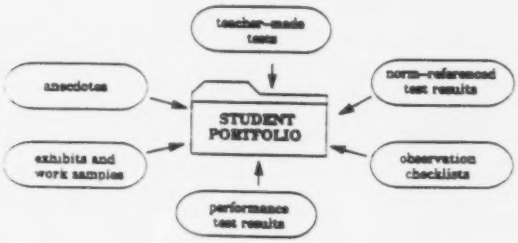


Figure 1

If we can clarify our terminology, we may be better able to deal with testing issues. The term "performance and product evaluation" used by Fitzpatrick and Morrison (1971) seems to describe the "new" measures accurately. Further, Fitzpatrick and Morrison provide a complete description of, and rationale for, performance testing that is as timely today as when it was written nearly 20 years ago. In fact, it is more timely today than in the past because performance testing has only recently captured the attention of the educational community. In the past, performance assessment had been primarily restricted to on-the-job performance evaluation.

In Chapter 8 of the Third Edition of Educational Measurement (Linn, 1989) Millman and Greene observe that "this single chapter has the responsibility for the topics found in seven chapters of the second edition...." (p. 335) Appendix A of this paper presents an abstract of one of the chapters to which they refer. The discussion of performance and product evaluation by Fitzpatrick and Morrison (1971) is directly relevant to the current interest in "authentic" academic assessment. The chapter by Fitzpatrick and Morrison not only provides an excellent overview of fundamental issues and concepts but it serves to remind us that performance testing is not a new phenomenon.

We will reiterate only two points from Fitzpatrick and Morrison because their discussion should be read in its entirety. First, the notion of performance testing implies the simulation of an actual situation, and, second, the degree of realism ("fidelity of simulation") can range from complete artificiality to the observation of the task as it is performed in the "real world." As performance testing becomes popular, we can expect to see performance tests whose fidelity of simulation ends at the title of the test and, at the other end of the spectrum, tests that are very realistic but not very practical or useful.

To some extent, any measure of performance becomes less than truly authentic when it becomes a performance measure because the process of observing the performance invariably influences the behavior of the performer. The desired degree of realism also raises cost/benefit considerations such as "How real is real enough?" and "How real can we get?" The illustration shown below suggests that complete realism is not always necessary or desirable.

A TRUE
PERFORMANCE TEST

Figure 2

The cost and/or consequences of accurate simulation of the criterion situation must be considered in deciding the best way to evaluate performance. During World War II, parachute packers were at times required to jump out of airplanes wearing randomly selected parachutes but this ultimate criterion measure was augmented by observation of skill in packing parachutes, questioning by supervisors, and, yes, even multiple-choice testing using items of the type shown below.

A PAPER-AND-PENCIL
PERFORMANCE TEST ITEM

1. How should you fold the parachute shown above before you place it in the backpack ?

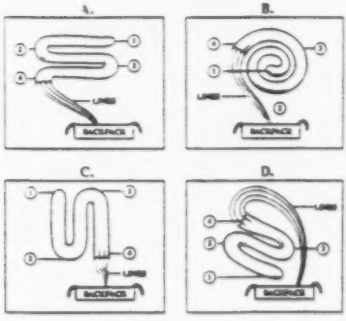


Figure 3

Educational performance tests have several distinguishing characteristics. They are designed to be as realistic as is practical, and they should contain no multiple-choice items. Computer processing of multiple-choice items is so tantalizingly easy and inexpensive that there will be an almost irresistible temptation to develop multiple-choice versions of educational performance tests. This would not be the first time that expediency has compromised the integrity of a great idea.

Performance tests are also focused, formative assessments which help teachers understand the processes students apply in solving problems. They are focused because they provide in-depth coverage of a fairly narrow band of exercises. They should be formative so that they direct student learning. And finally, they allow students to document the process by which they arrive at the solutions to the exercises contained in the assessment.

Figure 4 below contrasts the major characteristics of objective and performance tests.

Performance Tests (PT) vs. Objective Tests (OT)

PT: Student-Constructed Responses
OT: Student-Selected Responses

PT: Focuses on the Process of Problem Solving
OT: Attends to the Result of Problem Solving

PT: Scored by Teachers
OT: Scored by Computers

PT: Criterion Referenced
OT: Norm Referenced

Figure 4

One important point requiring discussion is that of measurement precision. In brief, portfolios, "authentic assessments," and/or performance tests are often substantially less objective and reliable than multiple-choice tests. This lack of precision results from the nature of the scoring process. Using portfolios and other performance measures to assess student performance is also much less efficient and much more expensive (by, perhaps, a factor of 10) than assessing performance using objective tests.

A benefit of using performance tests is that they more closely simulate the give-and-take of classroom instruction than multiple-choice tests. In addition, good educational performance tests can provide a model for classroom instruction and evaluation. They also provide new ideas for creative teaching. For these reasons, performance tests, judiciously used, can provide an additional dimension of information to norm-referenced test results and vice-versa.

One important point requiring discussion is that of measurement precision. In brief, portfolios, "authentic assessments," and/or performance tests are often substantially less objective and reliable than multiple-choice tests. This lack of precision results from the nature of the scoring process. Using portfolios and other performance measures to assess student performance is also much less efficient and much more expensive (by, perhaps, a factor of 10) than assessing performance using objective tests.

A benefit of using performance tests is that they more closely simulate the give-and-take of classroom instruction than multiple-choice tests. In addition, good educational performance tests can provide a model for classroom instruction and evaluation. They also provide new ideas for creative teaching. For these reasons, performance tests, judiciously used, can provide an additional dimension of information to norm-referenced test results and vice-versa.

To my knowledge, the performance tests developed by Riverside for the Arizona Department of Education represent the first major performance testing series that has been developed for assessing reading, writing, and mathematics. These tests were field-tested in spring 1990. There may be, however, other efforts under way that have not come to our attention.

No discussion of student performance testing should exclude student portfolios because the two topics have become intertwined in the literature. In discussing student portfolios, Valencia (1990) points out that student evaluation should be considered in view of "an expanded definition of assessment in which a wide variety of indicators of learning are gathered across many situations before, during, and after instruction. It is a philosophy that honors both the process and the products of learning as well as the active participation of the teacher and the students in their own evaluation and growth." (p. 340) As Valencia suggests, a student portfolio provides the means by which a variety of measures are brought together.

Evidence of Riverside's commitment to performance testing is documented by our participation in the Arizona program and by our new product, the Integrated Literature and Language Arts Portfolio Program. The two projects, Arizona and ILALAPP, resulted in a fortunate cross-pollination of ideas, which made both programs better than if they had been developed in isolation. In January 1991, we will publish a monograph entitled Educational Performance Assessment. This publication will contain chapters provided by several prominent researchers and practitioners and will provide an in-depth discussion of many issues associated with educational performance testing. It will also include the entire text of the Fitzpatrick and Morrison chapter referenced earlier in this paper. Reproduction of this chapter has been authorized by the copyright holder, the American Council on Education.

The law of the instrument suggests that multiple-choice tests are not a cure-all, but neither are performance measures, authentic assessments, or student portfolios. Standard student performance exercises can provide a new and useful dimension for student assessment, but the millennium will not arrive when they become popular, and those who assert that alternative assessments are a panacea for education are guilty of, at least, an excess of enthusiasm.

In the best of all possible worlds, educators will use free-response performance tests and multiple-choice tests selectively so that the most appropriate procedure is used for each specific measurement need.

References

American Educational Research Association, American Psychological Association and National Council on Measurement in Education, Standards for Educational and Psychological Testing, The American Psychological Association, 1985

Anastasi, Anne, Psychological Testing: Sixth Edition, Macmillan Publishing Company, 1988

Archibald, Doug A. and Newmann, Fred M., Beyond Standardized Testing, National Association of Secondary School Principals, 1988

Berk, Ronald A. (ed.), A Guide To Criterion-Referenced Test Construction, The Johns Hopkins University Press, 1984

Finch, F. L., "A Taxonomy for Competency Testing Programs," in Jaeger, Richard M., and Tittle, Carol Kehr (eds.), Minimum Competency Achievement Testing: Motives, Models, Measures, and Consequences, McCutchan Publishing Corporation, 1980

Finch, F. L., Toward a Definition for Educational Performance Assessments, paper presented at the Phi Delta Kappa/ERIC conference on "Alternative Assessments of Performance in the Language Arts," Bloomington, IN, August 1990

Fitzpatrick, Robert and Morrison, Edward J., "Performance and Product Evaluation," in Thorndike, R. L. (ed.), Educational Measurement: Second Edition, American Council on Education, 1970

Millman, Jason and Greene, Jennifer, "The Specification and Development of Tests of Achievement and Ability," in Linn, Robert L. (ed.), Educational Measurement: Third Edition, American Council on Education and Macmillan Publishing Company, 1989

Rothman, Robert, New Tests Based on Performance Raise Questions, Education Week, September 12, 1990

Valencia, Sheila, "A Portfolio Approach to Classroom Reading Assessment: The whys, Whats, and Hows," The Reading Teacher, January, 1990.

/So

APPENDIX A

The following has been abstracted from Educational Measurement, Second Edition by permission of the American Council on Education, Copyright 1971.

Performance and Product Evaluation (pp. 237-268)

Robert Fitzpatrick

Edward J. Morrison

It is apparent that there are many degrees and kinds of artificialities in tests. A test of the class here designated as **performance and product evaluation** is one in which some criterion situation is simulated to a much greater degree than is represented by the usual paper-and-pencil test. This type of test usually is called a **performance test**, and that term will be used here interchangeably with the more complete **performance and product evaluation**.

There is no absolute distinction between performance tests and other classes of tests--the performance test is one that is relatively realistic.

SIMULATION

The idea of **simulation** has been used in defining performance tests.

Fidelity of Simulation

The fidelity of a simulation, its degree of realism, presumably ranges along some scale from complete artificiality to the actual real-life situation.

Representativeness, Validity, and Reliability

The validity of a simulation means the degree to which it is relevant to its purpose. In the case of a performance test, validity is the degree of correspondence between performance on the test and ability to perform the criterion activity.

It is tempting to conclude that high representativeness of simulation and high test validity are synonymous. However, there are at least two problems with such a conclusion.

First, it must be remembered that fidelity is not a unidimensional quality and that a given amount of change in fidelity of one aspect may not have the same effect as a similar change in another aspect. Second, there may often be interactive effects such that a change toward higher fidelity in one aspect is accompanied by a change toward lesser fidelity in another aspect, or an increase in comprehensiveness may be gained at the cost of lower fidelity of certain aspects.

One further problem in performance tests is that of reliability. It is characteristic of real-life situations that they are difficult to control.

Good measurement is possible only when each examinee can be observed under similar circumstances; that is, when it it possible to control and hence standardize the displays, the surround, and the responses on which evaluation of performance will be based. Such control is characteristic of tests and is reflected in the high reliability of measurement that can be achieved with a good test.

But as the test situation simulates reality more closely, control becomes more difficult. It generally would be agreed by those with experience in the matter that the more closely one tries to simulate a real criterion situation, the less reliable will be one's measurement of performance.

TYPES OF TEST SITUATIONS

General situations, or contexts, in which performance tests may be cast are quite variable. No neat scheme has been proposed for classifying test situations; however, it may be useful to describe some types of situations as a means of illustrating the range of choices available in the development of performance tests.

Situational Tests

The term situational tests has come to be applied to tests in which the examinee is told to pretend that he is engaged in some real-life task, the nature and context of which is described to him in some detail before he begins to play his assigned role.

In-Basket Tests

A type of situational test important enough to be discussed separately is the in-basket test.

An in-basket test is a rather elaborate, realistic situational test intended to stimulate certain aspects of the job of an administrator. It consists of the letters, memoranda, records of in-coming telephone calls, and other materials that have supposedly collected in the in-basket of an administrative officer.

Work-Sample Tests

Perhaps the most common, if least commonly standardized, type of test is the work sample. The "boss" tells the examinee to perform a job-relevant task and assesses performance by observing either process or product.

Projects, Contests, and Rehearsed Performances

Special projects often are assigned by teachers for students to complete, usually out of class. Many such projects can be conceived as performance tests, and it may often be useful if the teacher so conceives them.

However, the evaluation of a project is made difficult by the fact that each student is usually doing a quite different project.

DEVELOPING PERFORMANCE TESTS

Studying Objectives

In the development of any test, it is of first importance to define carefully the objectives of measurement. In most cases, these follow directly from the educational objectives. Educational objectives specify the attainment of a given level of capability in certain criterion performances. When performance testing is appropriate, it is often necessary to have more than the usual amount of information about the criterion performances. This is so because it is necessary, in a performance test simulation of even moderately high fidelity and comprehensiveness, to know quite specifically what is to be simulated.

Formulating the Test Concept

If the criterion performance reasonably can be viewed as consisting of an aggregation of a large number of task or skill elements, one could simply enumerate the elements and choose at random for the test as many elements as necessary or feasible.

It is generally the best practice to include in the test several short tasks rather than one or two relatively long ones. Better comprehensiveness of coverage usually will be achieved in this way. Reliability of measurement also may be enhanced. Furthermore, this practice minimizes the likelihood that some examinees will have undue advantages in that they practiced a test task beforehand, either coincidentally or as a result of prior knowledge of the test content. However, the tasks chosen must be substantial enough to cover the important skills and to permit adequate simulation. Also, consideration should be given to efficiency in the use of testing time.

Stimulus Aspects

The stimulus aspects of the test situation may conveniently be described under three headings.

Instruction

It may often be possible through written or oral instructions to influence not only the examinee's specific responses to other test stimuli but also his attitudes and approaches to the test situation.

The instructions should teach each examinee how to take the test. In general, the test is not meant to be a measure of ability to take tests or to comprehend difficult instructions, but rather it is intended to get at the ability to perform in a certain way after that way is clearly understood. It may be necessary to devote a substantial portion of the testing time to this instructional function.

Displays

The displays present the stimuli to which the examinee is to respond. In a paper-and-pencil test, the displays are the test items. They are static and generally use relatively abstract symbols to represent or describe the real world.

In a performance test, in contrast, the displays need not be static and may be quite concretely realistic. The displays may be presented in writing, through speech, by diagrams, as numbers or formulas, or by any of a variety of other symbolic or physical means. Physical objects or equipment may serve as, or include, displays.

Surround

In any test situation, the surround should be arranged so that unwanted distractions are excluded and good performance is encouraged.

Other aspects of the surround are sometimes simulated merely as a means of helping the examinee to perceive the situation as a real one. If such features are inexpensive, there is likely to be no harm in them and they may sometimes do some good.

Careful analysis often is necessary to identify the important displays and to evaluate the influence of the surround under various conditions of performance.

Response fidelity

The examinee may be required, or afforded opportunity, to make responses which are more or less similar to those he would make in the criterion situation. If the response simulation is to be at high fidelity, it normally would be necessary for the examinee actually to respond in whatever ways are normal in the activity. At low fidelity, he typically would describe how he would respond, given particular stimuli.

Use of equipment

Many criterion performances require or allow the use of equipment through which the responses of the performer are mediated. Uniform policies about use of equipment should be established for all examinees.

Performance aids

Some kinds of equipment or materials are not directly involved in the performance but are such as to assist the performer.

Performance vs. Product

A fundamental question is whether to evaluate performance or product or both. Three issues need to be considered before some answers are suggested.

Nature of performance

The required performance may be such as to dictate or make academic the question of product vs. performance evaluation. Some performances have no product or, at any rate, performance and product are indistinguishable: examples of such performances are dancing and public speaking. In some cases, the process may be judged irrelevant, or there may be so many acceptable variations in performance that product alone is important; for example, a teacher of creative writing is unlikely to care how the student arrived at his product, short of plagiarism.

Feasibility of measurement

Sometimes, it is clear that the product is what ought to be measured, but it cannot be measured well enough. Or, the product could be measured adequately, but to do so would cost more than can be afforded.

Test security

A serious problem with some performance tests is that of security. Occasionally with intent but more often inadvertently, examinees may gain advance knowledge about a test and thereby obtain an unfair advantage over others.

Some specific advice on test security is given by Highland (1955):

1. Give as many tasks or units of activity to each individual as the available supply and practical considerations permit ...

2. If possible, devise and administer equivalent forms of the test so that a student never knows on which form he is to be tested ...

3. Arrange the situation so that those tested have no opportunity to talk to other class members not yet tested.

4. Emphasize that helping other class members will tend to lower the individual's own score.

5. If the test consists of several tasks or units of activity, change the order of administering these as frequently as possible.

6. Make frequent changes in the place within the testing room where a particular task or problem is given so that it isn't possible for examinees to pass along information to others not yet tested ...

7. Deny the examinee information concerning the rightness or wrongness of his response. This operates against the usefulness of the test as a learning device, but usually performance tests are administered because of their measurement rather than their teaching characteristics (pp. 54-55).

Observing and Recording Performance

If there is a more or less permanent product to be evaluated, it is normally not necessary to observe it until after it has been completed.

On the other hand, if the evaluation is of a performance or if the product is evanescent (e.g. writing on a blackboard), then it is necessary to observe and record relevant aspects of the performance as it occurs. It would in principle be possible to transform almost all performances into products by recording them on audio and video tapes or other media, and for some purposes it is desirable to do so. However, the expense of such a transformation would often be difficult to justify.

Scoring Performance Tests

The purpose of a score is to represent 91 concisely. A score is usually a number but may be another kind of symbol or verbal statement such as Pass or Fail.

USE OF PERFORMANCE TESTS

The potential value of the performance test lies in its closer approach to reality--its greater relevance in determining the degree to which the examinee can actually perform the tasks of the criterion job or other situation. This value generally is achieved at some cost, not only in money and other resources but in reliability of measurement. If the loss of reliability becomes extreme, the relevance is illusory since it cannot be measured. The test developer must weigh the factors of relevance, reliability, and cost in deciding how far in the direction of faithful simulation his test plans should go. The test user has a similar problem. There is no formula to help make these calculations.

If an adequately relevant and otherwise suitable paper-and-pencil test is available or can readily be developed, there is no point in using or developing a performance test. However, the ready availability of paper-and-pencil tests has often blinded us to considerations of relevance. It is an exercise in futility to measure accurately something one does not want to know. Relevance is the primary consideration, and good measurement is only a means to the end of appropriate evaluation.

mkk/37
7-6-89

TOWARD A DEFINITION FOR EDUCATIONAL PERFORMANCE ASSESSMENT

by Fredrick L. Finch

The Riverside Publishing Company
8420 West Bryn Mawr Avenue, Chicago, Illinois 60631

A paper presented at the ERIC/PDK Symposium

"Alternative Assessment of Performance in the Language Arts"
Bloomington, Indiana August 23, 1990

During the last twelve months it has been difficult to pick up a copy of Education Week or the Phi Delta Kappan without finding a reference to a need for "new" tests. I have read the manifesto of the "Campaign for Genuine Accountability" and the proceedings of the "Beyond the Bubble" conference. I have read Beyond Standardized Testing and about "authentic assessments" and "portfolio tests." All of this literature reminds me of one of my favorite verses in the Rubaiyat by Omar Khayyam:

> Myself when young did eagerly frequent
> Doctor and Saint, and heard great argument
> About it and about: but evermore
> Came out by the same door where in I went.

Each writer and speaker seems to have his or her own idea of what constitutes the so-called "better" tests that we should have. I think that the prize for obfuscation should be awarded to the International Reading Association speaker who stated that authentic tests are obviously superior to existing tests because they are more authentic. At the other end of the spectrum, I would award the prize for the best operational definition to Lois Easton of the Arizona Department of Education. She said that the Department wanted to develop performance assessments of state educational goals which did not rely on multiple-choice items.

Riverside has spent about eighteen months working with the Arizona Department of Education on performance tests designed to measure Arizona's Essential Skills. These tests will be used in conjunction with norm-referenced tests, the Iowa Tests of Basic Skills and the Tests of Achievement and Proficiency, to provide information about student performance as it relates to both national and state standards. The concepts and principles which evolved during the development of the Arizona performance tests were also applied to the development of Riverside's Integrated Literature and Language Arts Portfolio Program. We believe that these projects provide two examples of good educational performance assessment.

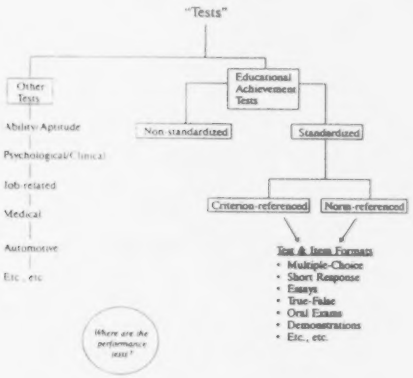
Before attempting to formulate a definition for educational performance assessment, I would like to briefly discuss some terms intrinsic to tests and testing.

COPYRIGHT © 1990 THE RIVERSIDE PUBLISHING COMPANY. ALL RIGHTS RESERVED.

Figure 1 shows a simplified family tree for tests. Since our primary interest is in educational achievement tests, I took the liberty of including all other tests under the classification "Other." For example, there are ability and aptitude tests, psychological and clinical tests, job-related tests, medical tests, automotive tests, and so on.

On the educational achievement test section of this chart, the first dichotomy that seems reasonable illustrates that there is a difference between standardized tests and non-standardized tests. This is particularly important in view of the fact that there seems to be some confusion about the term "standardized tests." Also, some have assumed that the terms "standardized" and "norm-referenced" are synonymous. This is not the case as indicated by the second dichotomy. Norms are developed during the process of standardization, but tests without norms may also be appropriately labeled "standardized tests."

A simplified family tree



Most critics of "standardized tests" also use this term in a restrictive manner. They equate standardized tests with multiple-choice items, whereas "standardized" means that the tests are administered under standard conditions. While it is true that most standardized tests use multiple-choice items, other standardized tests do not.

A standardized test is one that uses uniform procedures for administering and scoring the test (Anastasi, p. 25, 1988). Those who know the correct definition understand that this is not a small point. Without standard procedures for administering and scoring a test, it is not possible to compare scores earned by different persons. For this reason, non-standardized tests have limited use for any practical purpose. Consider, for example, a simple medical test such as taking a person's temperature. If the thermometer were not placed in a standard position for a standard amount of time, the temperature reading would be, for all intents and purposes, meaningless. If the correct terminology is used, there is little of interest "beyond standardized testing."

Standardized tests are of two primary types: criterion-referenced and norm-referenced. The distinction between the two is, I am sure, known to all. Either of these two types of tests may be based upon a wide range of item and testing formats, a few of which are

*Multiple-choice items
*Short-response items
*Essays
*True-false items
*Oral examinations
*Demonstrations.

This list could be expanded to include an almost endless variety of test and item formats.

Riverside's work on the Arizona project and in developing the Integrated Literature and Language Arts Portfolio Program leads us to believe that the primary distinction between performance tests and other tests occurs at the item level. As with any other useful test, a performance test must be administered under standard conditions and scored in a standard manner to provide consistency of test results. Performance tests are not beyond standardized testing; they should be standardized tests.

We have taken the position that performance tests can be most easily identified by exclusion. They are not multiple-choice tests. The distinguishing characteristic of a performance test is that measurement takes place in a fairly realistic situation. I think that most educators interested in alternatives to multiple-choice tests would agree that essay tests of writing are performance tests. Judging the competence of a painter, quarterback, or bassoon player could be an appropriate use of a performance test. In performance testing, the student constructs or provides the response rather than selects an answer choice.

A complete discussion of performance and product evaluation is provided by Robert Fitzpatrick and Edward J. Morrison in the second edition of Educational Measurement. Since this section was eliminated in the third edition, The Riverside Publishing Company has made arrangements with the American Council on Education to reprint this chapter in its entirety in the book Educational Performance Assessment, which will be published by Riverside in January 1991. This book will also include contributions by Lois Easton of the Arizona Department of Education, Morris Mueller from the Sacramento School District, Steven Osterlind of the University of Missouri, Nambury Raju of the Illinois Institute of Technology, Sheila Valencia from the University of Washington, and Lonnie Valentine from the Air Force Human Resources Laboratory.

Figure 2 provides a simplified overview of some of the characteristics of performance tests as contrasted with the characteristics of objective tests.

Objective Tests	Performance Tests
• Student-Selected Responses	• Student-Constructed Responses
• Attends to the *Result* of Problem Solving	• Focuses on the *Process* of Problem Solving
• Scored by Computers	• Scored by Teachers
• Norm-Referenced	• Criterion-Referenced

BEST COPY AVAILABLE

During a performance test, the student constructs a response. When an objective (multiple-choice) test is used, the student selects a response from a set of answer choices.

Performance tests focus on the process of problem solving. That is, performance tests help teachers to understand how the student arrived at a specific solution. In contrast, objective tests document the result of problem solving and provide no evidence about the procedure used by the student to arrive at the correct answer.

Performance tests will normally be scored by teachers for two reasons. First, performance test responses are not such that they can be processed by high-speed scanners and scoring machines. Second, the person most interested in determining how the student arrived at a specific answer is the classroom teacher.

While this next area is most subject to debate, performance tests are usually criterion-referenced instruments rather than norm-referenced instruments because the process of constructing a performance test assumes or requires that test developers have in mind criteria which the students are expected to meet.

Finally, educational performance tests are most useful when they are designed to provide a model for good classroom instruction.

Those whose primary interest is in replacing norm-referenced tests with performance tests may find that the nature, purpose, and uses of the two kinds of tests are so different that an either/or choice is not appropriate. Norm-referenced tests are so efficient, cost effective, and useful that they will continue to serve as the foundation for group achievement testing. Performance tests, while inefficient in a variety of ways, provide an important opportunity to add an entirely new dimension to the assessment of students. For these reasons, teachers, school administrators, researchers, and test developers must work together to ensure that this opportunity is implemented in a rational and professional manner.

References

Anastasi, Anne, Psychological Testing: Sixth Edition, Macmillan Publishing Company, 1988

Archbald, Doug A. and Newmann, Fred M., Beyond Standardized Testing, National Association of Secondary School Principals, 1988

Berk, Ronald A., A Guide To Criterion-Referenced Test Construction, The Johns Hopkins University Press, 1980

Fitzpatrick, Robert and Morrison, Edward J., "Performance and Product Evaluation," Educational Measurement: Second Edition, American Council on Education, 1970

BEST COPY AVAILABLE

Houghton Mifflin Company

One Beacon Street Boston Massachusetts 02108
(617) 725-5000 Cable HOUGHTON

THE PUBLISHER'S ROLE IN THE EDUCATION PROCESS

More than 100 years of experience developing instructional materials have given Houghton Mifflin Company a strong sense of the role of a responsible educational publisher. It is our goal to make substantive and unique contributions to the quality of education by shaping, developing, and distributing educational materials that enhance teacher effectiveness and student mastery. We consider ongoing service and support to the classroom teacher to be an integral part of our commitment to the teaching/learning process.

Houghton Mifflin publishes educational materials for all the major elementary, secondary, and college subject areas. These publications include student textbooks and workbooks; teachers' manuals; enrichment and remedial resources and instructional management systems; standardized tests and scoring services; and professional development materials for teachers and prospective teachers. In order to address a variety of educational needs, these publications take many forms, ranging from printed books to software for computer-assisted and computer-managed instruction, interactive databases of information, and training modules on videocassettes.

The Publishing Process at Houghton Mifflin

Instructional effectiveness is the measure of the quality of our educational materials. Through every stage of the publishing process, from planning to teacher support, we are committed to ensuring quality in these publications.

Learning is always the focus of educational publishing at Houghton Mifflin. Educators' experience, pedagogical research, and careful monitoring of student performance help the publisher to define and shape the instructional materials needed in the classroom. Because instructional materials are the tools of the teaching professional, every stage of the publishing process entails close interaction with educators.

Information Collection

Effective educational publications begin with the collection of vast quantities of information from educational research studies, task forces, conferences, decades of classroom experience with earlier publications, standardized tests, educators, and our own consultants. This information forms the basis for the development of an educational program.

Planning

Very few educational publications at Houghton Mifflin begin with an unsolicited manuscript. Initial planning determines the publication's content and instructional approach. This includes the scope of subject matter to be covered, the sequence of concept development, and the sequence of skill introduction, each of which must reflect the pedagogical plan.

The organization and physical format of the material are also decided at this point.

Selection of appropriate authors, editors, advisers, and reviewers to develop the publication is a crucial part of the planning process. Most publications are developed by a team specially selected to develop them. The members of this team are generally subject-matter specialists, who are former or current teachers drawn from schools and universities. Each member plays a critical part in the development of an effective educational publication.

Writing/Editing

Under the guidance of Houghton Mifflin editors, the authors provide content and implement the instructional approach. The editors' major responsibility in this collaborative procedure is to ensure that the manuscript or software program provides effective instructional support to a broad range of teachers and students. Experts in the field serve as advisers. They verify the content and help guide development. Reviewers evaluate the finished products to ensure that the materials meet their content and instructional objectives.

Classroom Testing

When the manuscript has been completed, it is field-tested by teachers to ensure instructional effectiveness under classroom conditions. This step permits discovery and resolution of potential problems before the material is published. Testing in classrooms continues after publication and this information is incorporated into the planning and content of future publications.

Design/Art

Development of the graphic design and illustrations for the publication is the final step. These elements reinforce the organization of the material and are essential to the publication's motivational effectiveness. The design must support the pedagogy and appeal to students if the program is to facilitate learning.

Educational Publications and Support Services

The publishing process results in textbooks, printed and electronic support materials, and extensive support services designed to help the teacher teach effectively. The success of this teaching/learning process is the focus of every component of the instructional program — from the textbook to additional resource material, classroom management systems, and professional development materials for teachers. More than 80 full-time Houghton Mifflin consultants provide ongoing individual assistance to elementary and secondary school teachers using our textbooks and materials in the classroom.

BEST COPY AVAILABLE

Ongoing Commitment to Investment

The publishing process is an ongoing cycle that normally spans more than five years from the information gathering stage to sale and service in the classroom. By the time students begin using the textbook, the information collection and planning phase is already well along for the next edition of the program.

Every textbook program requires investment by the publisher throughout each stage of the development cycle. Much of the net cash flow from operations must be reinfused to sustain momentum in the business. On average, educational publishers annually reinvest from $.40 to $.50 of every dollar of net cash realized into new and revised editions of textbook programs.

As one example, it is now estimated that the cost to develop an entirely new elementary school reading textbook program is about $20 million and as many as seven years of work. Within three years after that program is introduced to the market, a minor revision is necessary to keep it current with research and curriculum. This requires an additional investment of as much as $3 to $5 million. And within three more years, or six years after the original program was introduced, a major revision costing an additional $8 to $10 million is necessary to ensure the program is current and competitive.

This cycle of investment is typical for all textbook programs and, in fact, is becoming increasingly shorter with revisions occurring every two years rather than every three. The commitment to ongoing investment in textbook programs is essential to ensure that U. S. educators and the children they teach have current educational materials available. A commitment to excellence must also be backed by a commitment of financial resources.

Letter to Senator Pell from Mr. Veliotes

Association of American Publishers, Inc.,
Washington, DC, March 22, 1991.

The Hon. Claiborne Pell, *Chairman,*
Subcommittee on Education, Arts and Humanities
Committee on Labor and Human Resources,
648 Dirksen Senate Office Building,
Washington, DC.

Re: Supplemental Comments of the Association of American Publishers on "National Testing" Issues

Dear Mr. Chairman: During the March 7, 1991 hearing on the reauthorization of the Office of Educational Research and Improvement (OERI), the subcommittee heard testimony on the issues surrounding a national test. The subcommittee agreed to hold open the record until today for the submission of additional views on the issues. The Association of American Publishers ("AAP") appreciated the opportunity to testify at the hearing and submits this letter as a supplement to its original testimony of Burt Faldet.

As commercial publishers, members of AAP historically have been involved in meeting the needs of school districts and state education programs for various types of assessment instruments that have been used to assist in the measurement and evaluation of the Nation's school children. For the future, AAP stands ready to play an active role in the debate over national testing issues, so that commercial test publishers may contribute to a consensus solution for improving the Nation's educational system. We recognize our responsibility to be as creative as necessary to help our society meet these new formidable and exciting challenges. In that respect, AAP's Testing Committee has already met with Governor Romer and intends to continue those productive discussions. We also will open a dialogue with members of the business community to discuss how commercial publishers may work with business leaders to better understand and respond to the perceived educational problems associated with the complex issues surrounding assessment.

NO SINGLE NATIONAL TEST

AAP strongly opposes any single test to be used nationwide. However, AAP supports the concept of a nationwide assessment program, provided it meets specific criteria that, from the experiences of commercial publishers, are required to ensure that all universally-administered assessments (as distinct from assessments of student samples which may be intended for use primarily to monitor school systems and/or trends) will be used to improve student instruction and school curriculum. In this way, AAP also believes that more accountability can be brought to the assessment system. At the same time, AAP firmly believes that test diversity and competition should continue to be integral to any reform in order to assure that the best quality products are available. Publishers want to assure this subcommittee that the use of multiple tests can still provide uniform comparative information.

NATIONAL ASSESSMENT CRITERIA

1. There must be multiple testing instruments available. Different tests serve different purposes. Some tests are designed for diagnostic purposes, others for accountability reasons and still others may focus on monitoring progress. No one test, indeed no one type of test, can accomplish all of the diverse goals and objectives of this country's educational system. Nevertheless, as indicated above, multiple tests used in a coordinated fashion can still provide uniform comparative data.

A test should not be used for purpoees for which it was not designed, nor should a test become the sole basis for determining how a student is doing. Consequently, AAP urges that the first important criterion for a national assessment program must be adherence to the concept that a single national test not be used for purposes of determining progress toward achieving national education goals and standards. Additionally, varieties of assessment formats, from normed-referenced, standardized tests to performance-based assessments, have a role in the mix. An explanation of the various types of tests and formats was included in Mr. Faldet's earlier written testimony.

2. Each assessment must be demonstrated to be valid, reliable and bias-free. Scrutiny, criticism and debate about tests and the role they play in the lives of students and whether they are biased or exclusionary is not a new or novel development of the day. Those very issues have confronted the testing community, including devel-

opers, users, administrators and others, for years. As a result, voluntary *Standards for Educational and Psychological Testing* were adopted by the American Psychological Association (along with the American Educational Research Association and the National Council on Measurement in Education). The Standards set forth the professionally-accepted constructs and procedures to be followed in developing psychological tests, determining their reliability and validity, implementing their use and describing research results.

APA's Standards have been widely cited with approval by Congress, state legislatures, courts and researchers. The most recent endorsement occurred during the House consideration of H.R. 1285, the Higher Education Technical Amendments of 1991, between Chairman Ford and Congressman George Miller (page H. 1812, March 19, 1991). Under these Standards, developers and publishers of any assessment instrument, regardless of whether it is multiple-choice or performance-based, must present research on scientific validity (whether the instrument predicts what it is designed to predict), reliability (whether its results are repeatable) and fairness to protected groups (whether there is bias against any identifiable group of test takers).

3. Any national testing program must encourage open competition among all test developers, including commercial publishers. Regardless of the types of assessments to be included in a national program, commercial test publishers have an extraordinary background and expertise in producing instruments that comport with the APA's Standards and with the equally applicable *Code of Fair Testing Practices in Education*. The Code is primarily focused upon professionally-developed tests used in and by educational institutions. It is the commitment to extensive research and development by commercial test publishers that makes them second to none. Because these are private sector companies, they invest significant monies each year in conducting scientific research.

Similarly, this subcommittee should be aware that AAP members have been involved in offering performanced-based assessments (e.g., writing and listening assessments) for years. Some publishers are also offering portfolio assessments as supplements to current test batteries. Accordingly, commercial publishers have already demonstrated the initiative and responsiveness to the demands for performanced-based assessments. Any national program should seek to continue an open competition in the development and administration of assessment instruments to ensure that the best products are available to meet emerging educational needs.

4. A national assessment system must supplement, not supplant, existing state and local assessment. AAP feels strongly that control over the attainment of national goals and selection of assessments must remain with the state and local authorities. These entities are in the best position to make the necessary refinements to any national assessment program that will ensure that it is tailored to the needs of the local community which is being served. As several witnesses have previously testified before this subcommittee, this flexibility is important to preserving the best in the American educational tradition. A national assessment program should not replace state and local testing programs, which may serve other purposes, particularly diagnostic.

5. State and local entities must integrate the national assessment system with ongoing assessment. Although local control must reside with the state and local education agencies, they should have in place an explicit and comprehensive program to integrate and coordinate a national assessment system with on-going assessments. AAP submits that such an integration approach will provide a mechanism to ensure that all universally-administered assessments will be linked to the improvement of instruction.

6. The assessments must be administered and the results interpreted in accordance with the APA's Standards and the Code. Again, reliance on the adopted standards already in place must be assured, so that all aspects of the national assessment program can function together. Accordingly, it is important that test administration and test interpretation equally meet the high standards currently set forth by the APA's Standards and the Code. It makes no sense for the development of various assessments to be predicated on these standards and not have that action followed up by requiring those who administer the tests and who evaluate the test results to conform as well.

7. Much more emphasis must be given to assessment training. Further to the notion that test administration and interpretation must be as rigorous as the development of the national assessment program, AAP recommends that state and local education agencies should have in place a program to train teachers and administrators on the use and interpretation of assessments, including but not limited to the national assessment system. Research studies have shown that teachers do not

make full use of test results, simply because they have not received adequate training on these principles. Out of this point also derives the added benefit of improved accountability, since better informed teachers and school administrators will be better able to report the results of assessments to pupils and parents. In fact, the general public may also benefit from the greater awareness and understanding of the role of testing in our society.

KEY ISSUES FOR A CONSENSUS

Before any national assessment system is implemented, however, the AAP urges this subcommittee to pay particular attention to the need to reach a consensus on several key items that will be crucial to the success of the national program Such a consensus must involve the participation of members of all affected communities of interest, including experts in assessment and measurement. For its part, AAP is earnestly desirous of contributing manpower and resources to such an effort and we would welcome the opportunity to discuss ways for accomplishing this avenue for national consensus with this subcommittee or any of the groups that appeared at the hearing.

The following issues are those upon which agreement should be reached because they appear to AAP to be relevant to the development of a national assessment program that can and will achieve the goals and objectives advocated by the hearing witnesses:

1. The content, skills, objectives, and sequence to be taught and assessed, as well as on the interpretation of test results, e.g., attainment or achievement levels, pass/fail levels, college bound vs. technical, etc.

2. The purposes of a national assessment program, e.g., system accountability; monitoring progress in the attainment of national education goals; evaluating the results of the effectiveness of instruction; evaluating and improving curriculum; diagnosing individual strengths and weaknesses; or certifying the levels of achievement and ability upon ending secondary education. Subsidiary to what purposes are served, AAP feels compelled to raise the point that "high stakes" accountability testing may provide great leverage for changing the educational system, but it must be approached cautiously so that there are not distortions or unintended consequences and so that it will produce improvements in instruction and learning.

3. The voluntary or compulsory nature of participation in a national assessment program.

4. Who for the development, administration, scoring, and other costs of the program.

5. The scale of assessment, e.g., testing all students or a sample; testing annually or at intervals; testing all grades or certain grades; testing all students or excluding or providing separate tests or scores for some (e.g., non-public schools; Chapter 1; learning-disabled; ESL; etc.).

6. The frequency of administration and of revisions of the assessment to assure comparability of results over time, while encouraging a dynamic and diverse assessment program.

7. The security or accessibility of the assessments, i.e., will test items be secure or released to the public; will test items be protected by copyright; will test items be released to be imbedded in other assessment instruments.

8. The equating or scaling of different assessments, particularly assessments using different formats (e.g., multiple-choice or performance assessments), in accordance with the Standards. On this point, AAP notes the difficulty of the technical task of equating or calibrating different assessment instruments to an "anchor" test is not to be underestimated.

9. The choice of format of the assessments, (multiple-choice, performance, portfolio, or combination), is a function of a variety of factors, including cost, purpose of the test, timing, etc. Each format has advantages and each has disadvantages. It is not accurate to state that a performance or "authentic" assessment is the "best" assessment format without qualification.

On this last point, AAP is in extreme disagreement with those who advocate the need to completely remove the use of standardized, normed-referenced tests. That is akin to "throwing out the baby with the bath water." The use of multiple-choice tests has produced useful information on an objective, valid, nondiscriminatory basis. This subcommittee should not, any more than some educators and researchers, "shoot the messenger" when the message may not be to their liking.

DEVELOPMENT OF NATIONAL STANDARDS

Finally, AAP assumes that this subcommittee recognizes that the task of developing the instruments for use in a national assessment program is secondary to the task of establishing national education standards. Here again, it is the challenge for every group involved in the educational system to contribute to a consensus-building effort as to what every child should know in any particular subject area in any particular grade level. AAP hopes that this subcommittee will assure that this effort does not become the classic case of "putting the cart before the horse." It is premature to develop an assessment system until a consensus is reached about what is being assessed.

AAP believes that the participation of commercial publishers in this important task can be utilized at critical points in the formulation, development and delivery of appropriate assessment instruments and to ensure that instructional materials are developed and available. Even though this effort promises to be lengthy, perhaps even on the order of the 10-year program development outlined by Dr. Lauren Resnick, commercial publishers are committed to assuring that the national assessment program will be administered, and the results interpreted and used, in a standardized, unbiased way.

CONCLUSION

For all the above reasons, AAP respectfully suggests that the subcommittee's attention should focus first on the need for a consensus of education standards and then, secondarily, on the need for a national assessment program, rather than on a single national test. AAP members, who currently comprise the largest segment of the assessment delivery system in place in this country, are willing and able to respond to the emerging needs of a changing view of what multiple assessment instruments, and possibly their delivery techniques, will be demanded by the year 2000, just as they have done over the last quarter of a century, when numerous "new" assessments and vogue education reforms have required similar commitments of resources.

If further hearings are scheduled, we would welcome the chance to report back to this subcommittee on the progress of our efforts. In addition, we would be pleased to answer any questions this subcommittee may have based on these supplemental comments.

Respectfully submitted,

NICHOLAS A. VELIOTES
President

LETTER TO SENATOR PELL FROM SISTER SHEEHAN

UNITED STATES CATHOLIC CONFERENCE,
DEPARTMENT OF EDUCATION,
OFFICE OF THE SECRETARY,
Washington, DC, March 21, 1991.

The Hon. CLAIRBORNE PELL, *Chairman*
Subcommittee on Education, Arts and Humanities,
U.S. States Senate,
Washington, DC.

DEAR SENATOR PELL: As Secretary of Education for the United States Catholic Conference, I would like to take this opportunity to express the interest that our department has in the on-going deliberations of your subcommittee concerning the implementation of the National Goals for Education in general, but most specifically with regard to the question of a national test/assessment program.

The issue of standards and their assessment is a complex but important issue with many implications. There is a need to examine all of these possibilities carefully before taking any form of final action. One thing that I believe we want to avoid is a piece of legislation, along the lines of a single national test as opposed to a more flexible process of assessment. Such a single test might be viewed by some as a "quick fix" approach that could easily create more questions that it may hope to answer.

While I do not intend to speak for the total private school community, I believe that it is fair to say that we are greatly interested in these questions and have been trying to assess the implications on our students, programs and staffs. Representatives of the Council for American Private Education have been meeting with Arch-

bishop Francis Schulte of New Orleans, who serves on the President's Advisory Committee on the National Goals, and have expressed some initial reactions relating to the implementation of the goals and specifically our concerns in the area of assessment. I believe that in the future, representatives of the diverse elements of the private school community may wish the opportunity to address these issues in detail with your subcommittee.

The private school community in general, and the Catholic school community in particular, plays a significant role in the educational well being of our Nation. We have a strong interest in any proposal aimed at improving the education of our Nation's young people. I believe that we can offer a valuable contribution to any debate on how to attain this admirable goal and therefore I believe that the private school community should be included wherever possible, in such deliberations and on committees which may be formed to address basic issues of interest to the entire educational community.

I hope you will place my comments in the record of the subcommittee hearings. Please accept my thanks for your efforts on behalf of our school age young people.

Sincerely,

SISTER LOURDES SHEEHAN, R.S.M
Secretary of Education

[Due to the high cost of printing, the additional copy submitted by Sister Sheehan is retained in the files of the committee.]

PREPARED STATEMENT FROM FAIRTEST, NATIONAL CENTER FOR FAIR & OPEN TESTING

OPEN LETTER TO CONGRESS, BUSH ADMINISTRATION, THE GOVERNORS ON NAGB AND NAEP EXPANSION

Over the past several months, the National Assessment Governing Board (NAGB) has taken several actions which, considered together, raise serious concern over the future direction of the National Assessment of Educational Progress (NAEP). As a group of education and civil rights organizations active in school reform issues, we are addressing our concerns to Congress, the administration and the National Governors Association so that all responsible parties understand the nature of these problems and carefully monitor developments in NAGB and NAEP. It is important to note that we are not writing to oppose the national assessment, but to help ensure that it plays a constructive, not harmful, role in reforming our Nation's educational systems.

The actions of the governing board, taken together, go far beyond the level of activity authorized in the National Assessment of Educational Progress Improvement Act adopted as part of the Hawkins-Stafford Elementary and Secondary Education Amendments of 1988. That act (P.L. 100-297), which passed following lengthy discussion, authorized voluntary state-by-state comparisons of NAEP assessment results on a trial basis, and mandated an independent study of the validity and effects of the pilot programs.

Less than 2 years later, prior to completion of the trial comparisons and the studies, NAGB is proposing a major expansion of NAEP (see NAGB's paper "Positions on the Future of the National Assessment"). The proposal includes: (1) full participation by the states in state-by-state comparisons, to be paid for by the Federal Government; (2) testing and comparing local districts and even schools, which is currently prohibited by law; and (3) more frequent testing. Last month, NAGB adopted a process for setting "achievement levels" that students in grades 4, 8 and 12 ought to attain on NAEP tests (see NAGB paper, "Setting Appropriate Achievement Levels").

While each of these initiatives raises problems that require serious attention, we are particularly concerned about the combination of setting achievement levels and expanding NAEP. Our specific concerns and recommendations include:

(1) The proposal to expand NAEP was adopted before completion of the Congressionally-mandated studies or the pilot state-by-state comparisons.

Expansion of NAEP will inevitably affect our Nation's education. Congress correctly planned a cautious, step-by-step process to gauge the value and effects of state comparisons before mandating their continuation or expansion. This evaluation should be completed before any further steps are taken to expand NAEP.

(2) NAGB is proposing expansion of NAEP before the national debate on educational goals is resolved.

So far, the Bush administration and the governors have agreed on broad national goals, but have yet to decide how to implement them. Logically, the administration, the governors and Congress should all have roles in this debate as well as in determining the indicators used to measure progress toward the goals. But if measurement precedes goals clarification, the process of measuring becomes, by default, the process of defining. That would truly be putting the cart before the horse.

Deferring action on NAEP expansion until after the trial state comparisons and the legally required studies are completed will allow time for the national debate on attaining educational goals to reach resolution. Only then can NAEP play a constructive role in developing appropriate measurement tools and procedures.

(3) It is reckless to consider lifting the ban on district-by-district or school-by-school comparisons without considering the consequences for curriculum and instruction.

No one yet knows the effects—and side-effects—even of state-by-state comparisons. Repeal of the ban on local comparisons requires much more information and public discussion. It should not be considered until after the results of all trial comparisons and the mandated studies have been fully analyzed and publicly discussed.

(4) NAGB's achievement level setting process, when combined with comparisons, may create a de facto national curriculum.

The evidence is overwhelming that the more power attached to a test, the more control the test will have over curriculum and instruction. A national test with achievement goals and local comparisons will certainly become a powerful, perhaps controlling, influence on the curriculum.

The education goals enunciated by the administration and the governors do not attempt to mandate a national curriculum. In fact, there is widespread agreement that curriculum and instruction should not be determined from Washington. States and communities need flexibility in determining how to attain the broad goals. Yet NAGB's expansion proposals could preclude state and local initiatives.

(5) NAGB's achievement level setting procedures for its tests are not appropriate for determining national achievement goals.

The process chosen by NAGB to set achievement levels on its tests relies on selecting items from existing NAEP exams that, in the view of committees of experts, should be answered correctly by students who have attained the levels of "basic," "proficient" or "advanced." This is not an appropriate method for determining national curricular goals and achievement levels because it allows one test to define the content area and what students should be able to do in that area. Such decisions should be made prior to and independently of any test. After curricular goals have been decided at the various levels, then assessments appropriate to the curriculum can be constructed and achievement levels set.

Moreover, as the recently-released report of the National Commission on Testing and Public Policy explains, the procedure of relying on committees of experts to examine items is flawed even for the purpose of setting cut-off scores on tests. NAGB thus expects a limited technical procedure to be adequate for shaping a national curriculum.

(6) By setting achievement goals based on what are predominantly multiple-choice tests, NAGB runs the risk of defining national educational goals in terms dictated by these narrow instruments.

In potentially shaping curriculum and instruction, NAEP tests will affect both content and methods of teaching. Multiple-choice testing necessarily focuses on factual recall and simple comparisons and observations. It does not lend itself to revealing whether students know how to do something—to write a persuasive essay, research an historical event, or grasp the meaning of a scientific development.

The narrowness of these instruments has been recognized by the governors, among many others, and has led to widespread efforts to develop and implement other means of assessment. If multiple choice testing continues to predominate, NAEP will provide a continual obstacle to teaching and assessing the important things students need to learn how to do. It will help perpetuate a reduced definition of the content to be studied and an entirely incorrect view of how students learn.

(7) NAGB proposes to vastly increase the amount of its testing to include "at least three subjects each year."

The current NAEP authorization establishes a 2-year testing cycle and a minimum frequency for testing various subjects. Only math and reading are to be tested every 2 years; other subjects are scheduled at 4- or 6-year intervals. Though its futures paper deferred discussion of the "exact configuration" of the new testing cycles, NAGB called for "testing at least three subjects each year," at least six tests every 2 years. NAGB's claims this acceleration is necessary "to provide timely and

sufficient data" and to "replace the Education Department's annual 'wall chart' which relies on SAT and ACT scores."

Again, major changes in NAEP such as expanding the extent and frequency of testing should not be undertaken prior to completion and analysis of the 1992 testing and the mandated studies. In fact, such expansion is not at all necessary. Because educational systems and achievement cannot change rapidly, yearly aggregated data will not provide meaningful information about important educational changes. Less frequent information should be quite sufficient.

While virtually everyone, including Secretary Cavazos, agrees on the inadequacy of the current "wall charts," the mere existence of the charts is an insufficient justification for vastly increasing a national testing program. To be sure, annual one-point changes in average SAT scores or two-tenths of a point changes on the act in the "wall charts" are meaningless. But substituting minute changes in NAEP scores would not be an improvement. It could, however, produce public frustration and thereby jeopardize public support for educational reform. Maintaining NAEP's current, authorized schedule will provide as much useful information at less cost in dollars and, ultimately, in public credibility.

(8) NAGB is moving too slowly in revising NAEP exams to rely less on multiple-choice questions and to develop other means of assessment which better measure the full range of knowledge and skills.

While NAGB claims that about 20 percent of this year's NAEP math items were open-ended, Paul LeMahieu, Pittsburgh's Director of Testing, informed the National Association of Test Directors that less than 5 percent were really open-ended items. The rest were multiple-choice questions with the answer options deleted. Like multiple-choice items, such questions are not very useful in measuring student abilities to use math to solve real-world problems.

Instead of expanding the use of outdated, multiple-choice tests, NAEP should become a leader in the national effort to develop improved forms of assessment that provide more information and do not endanger but rather enrich the curriculum. NAEP should work with the states, a number of which already have performance-based assessment projects under development, to produce and evaluate such assessments.

(9) NAEP expansion will absorb an ever larger share of Federal research and information dollars, but the results may not be worth the money.

The NAEP Improvement Act authorized $9,500,000 for fiscal year 1989 for NAEP. For fiscal year 1990, NAEP received $17,084,000. Even with this increased amount, the Education Department deferred the NAEP validity study, a national assessment of adult illiteracy and work on the National Education Longitudinal Study. For fiscal year 1991, NAGB has requested $18,866,000, an increase of more than 10 percent over fiscal year 1990 and nearly double the authorization for fiscal year 1989. NAGB receives up to 10 percent of NAEP funds for administrative purposes and reportedly seeks to receive up to 15 percent. Estimates of the cost of NAEP if expanded are $100 million annually, a more than five-fold increase over current expenditures and an amount two and one-half times the funding for the National Center for Educational Statistics (NCES).

Will the results be worth the additional money? Yearly testing will not increase anyone's knowledge of the effects of educational reform efforts. Further state and local comparisons may not tell us more than we already know about how well the states and localities perform on standardized tests. In a period of continuing fiscal restraint, money used for more extensive testing could be better used to improve the quality of NAEP assessments or for other needed research rather than for redundant and potentially dangerous increases in testing.

(10) The relationship among NAEP, NAGB and NCES must be clarified.

The current debates over the future of NAEP have raised questions about the appropriateness of an independent body wielding the power that NAGB could assert over our Nation's education. A key issue is whether such a body is adequately accountable to Congress, the administration and the public.

Since accountability is, in part, asserted by control over funding, NAGB's budget should be separated from NAEP's. So long as NAGB obtains a percentage of a (potentially rapidly-expanding) NAEP budget, there is no way for elected officials to adequately exert oversight. The role of NAGB in relation to NCES, the Department of Education or any other bodies created to oversee progress toward national goals should be carefully considered by the appropriate House and Senate committees and the administration before NAEP is expanded.

In sum, NAGB's plans to rapidly expand NAEP without adequate consideration of the effects of the expansion or the proper role of assessment in educational reform are dangerous. Neither Congress nor the administration should allow them to pro-

ceed without careful review and consideration. Similarly, the governors should not support the use of NAEP for measuring progress toward national goals without first clarifying the goals and the role of assessment in achieving them and then determining the details of measurement. Specifically:

—NAEP should not be expanded to allow more frequent or extensive testing or more detailed comparisons at least until completion of the trial assessments of 1990 and 1992 and the independent evaluation mandated in the act. Then, Congress, the administration and the governors must weigh carefully the potentially harmful effects of more extensive testing and comparisons and ascertain that the dangers do not outweigh any possible benefits. In any event, expansion of NAEP must be subsequent and subordinate to the establishment of national goals and not allowed to dictate a national curriculum.

—NAEP should be directed to spend a significant portion of its budget on developing and piloting performance-based assessments (including tests and portfolios). Such research and development should be planned carefully to coordinate with state projects such as those underway in California, Connecticut and Vermont, to develop performance-based assessments, as well as projects undertaken by local education authorities or other governmental or private bodies.

—Congress and the administration should consider separating NAGB funding from NAEP funding and carefully consider the future role of NAGB in relation to other agencies and bodies.

We appreciate your attention to these most important issues and look forward to working with you in the effort to achieve genuine and lasting reforms in the quality of public education.

Please feel free to call any of us if you have any questions or need further information.

LIST OF SIGNERS

Ronald J. Abate, College of Education, Cleveland State University*
Advocates for Children of New York, Norman Rollins, Executive Director
American Association of Colleges for Teacher Education, David G. Imig, Executive Director
American Association of School Administrators
American Reading Council, Julia Palmer, President
Angelo N. Ancheta, Asian Law Alliance*
APPLE Corps., Inc., Alfred E. McWilliams, President; Mary Anne F. Gaunt, Executive Director
Association for Supervision and Curriculum Development, Gordon Cawelti, Executive Director
Bank Street College of Education, Joseph Shenker, President
Rims Barber, Mississippi Human Services Agenda
Leonard Beckum, Dean, College of Education, City College of New York
Association for Women in Science, Stephanie J. Bird, President
Gerald W. Bracey, Director of Research and Evaluation, Cherry Creek, Colorado
Diana Caballero, Puerto Rican/Latino Education Roundtable, New York City*
Center for Law and Social Policy, Alan W. Houseman, Director
Center for Women Policy Studies, Leslie R. Wolfe, Executive Director
Harold E. Dent, Vice-President, Psychological and Human Resources Consultants, Inc.*
Educational Law Center, Inc., Marilyn Morheuser, Executive Director
Pablo Eisenberg, President, Center for Community Change*
Federation of Organizations for Professional Women
Foxfire Fund, Inc., Eliot Wigginton, President
Fund for the Feminist Majority, Eleanor Smeal, President
Howard Gardner, Project Zero, Harvard University*
Genesee Valley Developmental Learning Group, New York
Leslie A. Hart, Brain-Compatible Education Associates
La Donna Harris, Americans for Indian Opportunity*
Asa Hilliard, Professor of Education, Georgia State University*
Holt Associates/Growing Without Schooling
Institute for Learning and Teaching, Wayne Jennings, Director
International Reading Association
KEY-Kids, Education and You, (Beth Bradley, Jenny Coston, Leslie Floyd, Sue Long)
A. Gay Kingman, National Congress of American Indians*

Nancy K. Klein, College of Education, Cleveland State University*
Massachusetts Advocacy Center, Stephen R. Bing, Executive Director
Matsushita Foundation, Inc., Dr. Sophie Sa, Executive Director
Deborah Meier, Principal, Central Park East Secondary School, New York City
Sara E. Melendez, Vice-Provost, University of Bridgeport*
Susan Metz, Prospect Heights High School, Brooklyn, NY*
Mexican American Women's National Association
National Association for the Advancement of Colored People, Beverly Cole, Education Director
National Association of Secondary School Principals
National Center for Fair & Open Testing (FairTest), Cinthia Schuman, Executive Director
National Coalition of Advocates for Students
National Coalition of Title I/Chapter I Parents, Robert Witherspoon, Director
National Council for the Social Studies
National Council of Teachers of English, Charles Suhor, Executive Director
National Education Association, Keith Geiger, President
National Indian Youth Council, Inc., Cheryl J. Mann, Executive Director
National Organization for Women—New York City Chapter
National Parent Teacher Association
National Women's Law Center
NationalWomen's Political Caucus
Fred M. Newmann, National Center on Effective Secondary Schools—University of Wisconsin*
New York Public Interest Research Group
NOW Legal Defense and Education Fund, Helen Neuborae, Executive Director
Organization of Chinese American Women, Faith Lee Breen, Chair, Board of Directors
Vito Perrone, Harvard University—Graduate School of Education*
Project Equality, Inc., Kansas City, MO
Puerto Rican Legal Defense and Education Fund, Inc., Ruben Franco, President/General Counsel
Rochester (NY) Teachers Association (AFF), Adam Urbanski, President
Lori Rubenstein, Partnership for Democracy*
William V. Schipper, Executive Director, Natl. Assn. of State Directors of Special Education*
Donald H. Smith, Chairman, Dept. of Education, Baruch College, City University of New York*
Hilton Smith, Coordinator, Foxfire Teacher Outreach Programs*
Southern Association on Children Under Six, Cathy Grace, Executive Director
Southern Christian Leadership Conference, Joseph Lowery, President
Southern Regional Council, Inc., Steve Suitts, Executive Director
Judy I. Stahlman, Cleveland State University*
Adria Steinberg, Harvard Education Letter*
Gail E. Thomas, Texas A & M University*
United States Student Association, Julianne Marley, President
Dorothy J. Watson, President, Whole Language Umbrella*
Paul Weckstein, Center for Law and Education*
Grant Wiggins, Center on Learning, Assessment and School Structures (CLASS)
Arthur E. Wise, Rand Corporation*
Women's Research and Education Institute, Betty Parsons Dooley, Executive Director

*Organization for identification purposes only

At this point, we adjourn the committee.
[Whereupon, at 12:30 p.m., the subcommittee was adjourned.]

BEST COPY AVAILABLE

REAUTHORIZATION OF THE OFFICE OF EDUCATIONAL RESEARCH AND IMPROVEMENT ACT

THURSDAY, JUNE 13, 1991

U.S. SENATE,
SUBCOMMITTEE ON EDUCATION, ARTS AND HUMANITIES, OF THE
COMMITTEE ON LABOR AND HUMAN RESOURCES,
Washington, DC.

The subcommittee met, pursuant to notice, at 10:04 a.m., in room SD-430, Dirksen Senate Office Building, Senator Claiborne Pell (chairman of the subcommittee) presiding.

Present: Senators Pell, Simon, Bingaman, Kassebaum, and Jeffords.

OPENING STATEMENT OF SENATOR PELL

Senator PELL. The subcommittee will come to order.

This marks the second hearing on the reauthorization of the Office of Educational Research and Improvement, OERI.

Our first hearing in March focused on the question of the national test. Today we will examine broader issues relating to educational research and improvement.

Earlier this week Senators Kassebaum, Kennedy, Hatch and I introduced S. 1275, a simple extension of OERI. This bill will serve as the vehicle for our substantive work, and as we look toward reauthorization, I think we all have several goals in mind.

The first is that we make the Office of Educational Research and Improvement a true beacon of educational excellence. The work of that office should be on the cutting edge of new, innovative and controversial educational ideas, concepts and programs.

The second is that dissemination and access be the primary mission of every component of the office. Educational research should be brought out of the hallowed halls of ivy-covered buildings, into the front lines of instruction in our Nation's schools.

The third is that we should make every effort to ensure that research remains free of any ideology. The research agenda of the office should have only one concern in mind—improving the educational achievement of every student in our Nation.

To have a critical impact upon teaching and learning, research must remain free from the influence of the tenets of any special interests.

Fourth, the office should serve as a conduit of successful school practice, and we should charge that office with the mission of identifying proven models of educational excellence and developing a

42-921 0 - 91 - 7

mechanism for transmitting that excellence into our Nation's schools.

Finally, the Secretary through the office should develop a national test or series of tests to measure the achievement of individual students. I have already introduced legislation that directs the Secretary either to develop or identify such tests, and it is our intent that we act on that legislation within the context of the present reauthorization.

We've got a golden opportunity in this reauthorization to make the office the real engine that drives the train of excellence and innovation. If we take this challenge seriously and responsibly, the results of our efforts will have a lasting imprint on educational excellence and improvement for years to come.

We welcome your statement, Mr. Secretary, and we have two panels which will follow you. First, I'll turn to Senator Bingaman. Do you have a statement you wish to make at this time, Senator Bingaman?

Senator BINGAMAN. I have no statement, Mr. Chairman.

Senator PELL. Thank you.

Secretary Alexander.

STATEMENT OF HON. LAMAR ALEXANDER, SECRETARY, U.S. DEPARTMENT OF EDUCATION, WASHINGTON, DC, ACCOMPANIED BY BRUNO MANNO, ACTING ASSISTANT SECRETARY, OFFICE OF EDUCATIONAL RESEARCH AND IMPROVEMENT

Secretary ALEXANDER. Mr. Chairman, Senator Bingaman, thank you for inviting me. I have submitted my statement; I'd like to make some brief comments on it, and then I'll be glad to try to answer any questions that you might have.

I'd like to introduce Dr. Bruno Manno, who is the acting assistant secretary for the Office of Educational Research and Improvement.

I will have a word or two to say, Senator, about the idea of the national examination. It is a subject that you have been a leader on for a number of years. In fact, I think many people are not aware that you have already persuaded the Congress to pass a law requiring a sort of national achievement examination some years ago.

The President has now suggested a voluntary national system of examination. His advisory committee on education, which is a broadly based group of educators and business leaders in America, has said the same thing. Many members of the National Governors Association have said the same thing. So it is an idea that is coming right along.

I received a letter from a 4th grade student in the Richardson Independent School District in Dallas, TX who is in a class with my nephew, Jeremy Carl, a few weeks ago, and she summed up, as 4th graders have a way of summing up almost the entire debate about the national examination system. She said, "Dear Secretary Alexander, having a national exam is both a good and a bad idea. It is a good idea because we can see what we need to learn and what we already know. It is a bad idea because it may not make

any difference, and then we would have wasted a lot of time. Love, Angela."

So I think she has come pretty close to summing up both sides of the argument.

I am here today, Mr. Chairman, to talk about the reauthorization of the Office of Educational Research and Improvement. It is a pleasure to be here to talk about that because that is one of the areas of activity of the United States Department of Education about which everyone seems to agree that this is what we ought to be doing. But sometimes we have an argument about what is Federal and what is local and what is State and whether we ought to be involved in this or that. But for more than a century it has been generally agreed that it is an appropriate function of the Federal Government to gather the best information about what is going on in education in America, about what works, to conduct research about that and to disseminate it as widely as possible to our 110,000 public and private schools in America to help improve the quality of education.

What we are talking about today, it seems to me, is gathering that information in a way that can most effectively help us close what might be called a skills and knowledge gap. And rather than talk about it in an abstract form, I think it is important to mention this work in some examples.

For example, just last week the National Center for Education Statistics brought out the two-volume work that it does each year on the condition of education. It has some fascinating information in there, useful to parents, to teachers, to members of the United States Congress. It shows, for example, that the high school completion rate for 19 year-olds in 1989 was 81 percent. It has some interesting information about the cost of going to college that contravenes some conventional wisdom and that would be very useful to members of this committee as we talk about what do we do about the Pell grants, what do we do about the loans; do we agree with the administration proposal to focus more of the money on the poorest Americans, or do we spread it out more among middle-income Americans.

What this shows, for example, in general, is that the cost of going to public universities, which is where most people go, for most American families has not increased in the last 25 years as a percent of their income; that the families for whom the cost has increased dramatically have been low-income Americans, which is one reason why the administration is seeking to focus more of the money we have on low-income Americans. We can debate that, but it helps to have reliable information that helps us make decisions.

We saw another good example of that last week with the work of the Nation's Report Card, the National Assessment of Education Progress, coming out with the first ever State by State date in what 8th graders know about mathematics. That was work authorized by this subcommittee and full committee and Congress a few years ago, and it was very helpful and very disturbing data. It showed, for example, that there is a wide variation of what 8th graders know about mathematics and can do about mathematics among the States. What is more important, it shows that none of the States are cutting it; that even in States where students score

the highest, only about one in four seem to know what most educators think an 8th grader ought to know and be able to do about mathematics.

That is very helpful information. It ought to ring an alarm bell across this country and drive a lot of the work community by community that will have to be done to help the Nation reach the national education goals by the year 2000.

That same work done by the Nation's Report Card gave interesting information about the effect of television watching on the learning of mathematics. It showed fundamentally that the children who watched the most television did the worst in math, and the children who watched the least did the best. It suggested that common sense things that we would know without looking do make real sense even when we research it—not only television watching, but the amount of homework completed, the amount of absenteeism at school, the disorganization of the family from which students come all seem to have something to do with how much math you learn. It even had interesting information about what students think they know about math as compared with what they do know and found that a great many American children who think they are good in math in fact don't know much math. And I suppose scholars will be studying that for a while. It shows we are high in self-esteem and low in knowledge in some respect.

The reason I mention those is so people won't think we are talking about abstract notions here that are not useful and available to broad numbers of Americans. In fact, this information is available on a daily basis and useful on a daily basis, not just to Members of Congress but to parents, school leaders, and decisionmakers everywhere.

One of the most important points I would like to make today is the following; that the work the committee will be doing this year and next year in reauthorizing and considering the reauthorization of OERI, this work is likely to take a year or two, and then it is likely to set the rules for what is done for the next 5 years after that. In other words, what we are talking about today will govern the amount of research and the kind of research and the contribution that the Federal Government can make to helping move the country toward the national education goals for the entire next decade because this reauthorization bill, when approved, will expire toward the end of this century, so we have to think pretty far ahead.

I would like to make as strong an argument as I respectfully can make, therefore, that the committee continue to give OERI as much flexibility as possible because the committee may want and the country may need for us to make decisions in 1994 and 1995 and 1996 that will be different to anticipate with precision today. We want to stay within what Congress wants us to do, but the tradition of a certain amount of flexibility within what you allow OERI to do I would submit is very important in a rapidly change world when the stakes are so high in terms of what we need to know and are able to do.

Arguably, our biggest problem is that so many people in America think the Nation is at risk but that they are okay, when they are not okay. And using this information to help ring and alarm bell

and wake the country up to what needs to be done is vitally important. That flexibility is an important part of doing that.

Within the next few weeks we will be sending to this committee and to Congress, Mr. Chairman, the specific terms of our legislation. Today I would simply like to review very briefly four aspects of the proposal that will be coming.

First, we'll be including in that the proposals that were made in the "America 2000: Excellence in Education Act". These would build on the work that Congress approved a couple, 3 years ago. Instead of just having 8th grade math tests State by State, we will be recommending that Congress approve the collection of State-level data in grades 4, 8, and 12 in the five core subjects, reading, writing, math, science and history, beginning in 1994.

We think parents need to know in their schools and in their communities the same kind of information that States found out last week on a State by State basis.

We will be recommending that the Federal Government pay for costs in excess of $100,000 associated with the administration of those assessments.

We will be recommending that we collect this data annually and that we permit the use of NAEP-like tests at the district and school levels by States that wish to do so.

I think the value of these exams has already been shown. We noticed last week that already, based upon the State by State reports, that whole States were calling for major reviews of the way they teach math and what they are teaching in math. The District of Columbia, we read about, is a smaller area, and the school officials there, based upon the information they receive, went to work immediately to see what sort of improvements they could make in the teaching and learning of mathematics.

I'd like to underscore a point about our request that we begin this collection of data in the five core subjects so that parents and communities can have it in 1994. If the Nation's Report Card, NAEP, is going to be able to prepare for 1994 assessments, we would need to ask that the Congress approve this by the end of this year. That would be a little in advance of what I would guess you would plan to do in terms of the entire reauthorization of OERI. But I would like to point out that timing problem if you would like for NAEP to be able to proceed with the 1994 assessments.

Mr. Chairman, you have mentioned in your remarks the idea of a national system of examination. There is not in this proposal specific authorization language that has to do with a voluntary national examination system that President Bush has talked about or the kind of examination system that you have talked about. I think it is important to say, however, that we are proceeding to do some of the preliminary work that would lead toward the objective that you outline in the legislation that you described.

For example, Congress has already authorized the U.S. Department of Education to do research work that will promote student achievement. Under that authority we are already providing technical assistance to such organizations as the National Governors Association as they do some preliminary work in thinking about what kind of national examination system would be most appropriate and how to go about doing it.

In addition the Congress has already passed a law, as we mentioned earlier, which you authorized which requires us, or at least gives us the option, to go ahead and develop a national examination. So for both of those reasons there is a good amount of work going on giving technical advice and thinking about a national examination system.

In addition to that, members of the Senate and members of the House have been working with the governors and with the administration to create an interim council on standards and testing which would work between now and the end of the year, with a broad-based group of educators, administration officials and members of Congress, and think primarily about how to proceed in terms of developing a national examination system but also to give some thought to the pros and cons of such a system.

It would be my suggestion that as this discussion moves along during the next few months—after all, 6 months isn't a very long period of time—that as this moves along, it will become clearer what kind of legislation would be the most appropriate next step for Congress to take. I am very pleased that members of this committee have agreed to participate in this discussion with the governors and with members of the administration to try to take these next steps.

Fundamentally, the President and you seem to agree on the need for a national examination system. The details of such a system are something that need to be thought about very carefully. There are many ways to go about such a system. The President is not thinking about a standardized test; he is not thinking about a multiple choice test; he is not thinking necessarily about a test written in Washington even, and he is not thinking about a test that every student in America would be required to take. He is more inclined to think of an examination system in the five core subjects—English, math, science, history and geography—that would be carefully developed over the next several years and that would be useful in helping to point the way toward a higher standard of learning and a higher standard of achievement that would be available to communities and available to parents, and the hope would be that they would adopt that examination or they would improve the examinations they are now using to reach at least the same high standards.

We can talk more about the ways to go about doing that if you would like to, but there is a substantial area of agreement here between the President's views, Mr. Chairman, and yours, and we would salute your leadership in the area and look forward to working with you in the development of an appropriate national examination system.

The other major areas that our legislation will cover have to do with increasing the flexibility of the way we spend some of our research dollars. There are a number of mandated minimums that say exactly how much the department should spend on this research center, on that regional laboratory, on this study or on that clearinghouse, and what we would like to suggest to the committee is that to try to anticipate that in 1991 to take effect all the way through 1998 is very difficult to do and that the more flexibility you can accord the department in making those decisions, the more useful the taxpayers' dollars can be spent.

We are always available to be hauled up here for an oversight hearing. We are always willing to accept strong suggestions from members of the Senate about what ought to be done or ought not to be done, and we would like to work cooperatively with you on that and not have either of our hands tied by laws that we might establish this year to have effect in 1997 and 1998.

Just a couple of other points. We would like to expand the types of entities that are eligible to compete for regional laboratory and research center awards. We believe that will help unleash the creativity of America more if we are not as limited in the way we can grant those dollars, which are getting to be a substantial number of dollars.

A third area that we would like to address and we will address when our legislation comes up is to take the current list of OERI priorities which have been expanded so much that they are more like a laundry list now than a list of priorities, and focus them better and cause the Congress to concentrate on trying to focus more on these.

Finally, we would like to do a better job of a challenge that this activity as a part of the Department of Education has always had—we'd like to do a better job of getting the research into practice.

I am hearing every day from people who say we'd like to know more about what works. One of the great frustrations in America in education is that people in a school that is not functioning very well very rarely will walk across the street to a school that is functioning well and borrow a good idea. It is hard to know why that doesn't happen, but one of our jobs is to see that it happens more frequently, and we'll be looking for ways to do that. One way to do it might be to give the national diffusion network, which is the name of a network whose job it is to help do that, a name that everybody could understand so they would know that it was doing. And we'll be giving some thought to that, too.

Fundamentally, I'd like to close where I began. There are examples all around us. There is an editorial in the *Washington Post* this morning about the work that the Center for Education Statistics is doing in helping us to understand the number of dropouts; the work done on the condition of education 1991 that helps us understand, among other things, the cost of going to college; the work reported last week, helping us understand what 8th graders know and can do about math State by State.

This is the kind of information that we are talking about in this reauthorization. We think it is extremely important. The advent of the proposed American achievement test or voluntary national system of examination is one of the most fundamental and important changes that would take place in American education in the two centuries that we have been a country. All of these are what we are talking about as we talk about reauthorizing OERI, and we look forward to working with members of the Senate and doing that reauthorization well.

Thank you.

[The prepared statement of Secretary Alexander follows:]

Prepared Statement of Secretary Alexander

Mr. Chairman, and Members of the Subcommittee, I am pleased to appear before you today to talk about reauthorization of the Office of Educational Research and Improvement (OERI). Carrying out OERI's function—providing reliable information about the condition of education and the improvement of learning—is one of the most important ways in which the Department can help move the Nation toward its national education goals.

Two weeks ago, the Department released a new book, the report of the National Center for Education Statistics entitled *The Condition of Education 1991.* From just one of the nearly 120 indicators included in the report, we learned that, as a Nation, we face a considerable challenge in reaching the national goal of a 90 percent graduation rate by the year 2000. Although a number of former dropouts complete high school by age 24, either by returning to school or through some alternative means such as the GED, the high school completion rate for 19-year-olds in 1989 was only 81 percent.

One week ago today, the National Center released information on the state of mathematics achievement from The Nation's Report Card. This report, which includes the first-ever State- by-State data from the National Assessment of Educational Progress, shows us the challenge we face in reaching our national goals related to students' math achievement. We know from the report that the challenge is great, that only five percent of U.S. twelfth graders have the knowledge and skills to be able to handle college level mathematics. We know that some States face a greater challenge than others, that the average proficiency of eighth graders in mathematics varied across the States by as much as 50 points, but we also know that in none of the States do students know enough math. What we don't know is what parents, teachers, and communities need to know, which is, "How well are my students and my schools performing?"

I'll return to that later. My point for the moment is that after these reports came out, everyone in the country knew something about where the Nation stands in relation to some of our education goals. We can thank the Office of Educational Research and Improvement, and specifically, the National Center for Education Statistics, for that. There has been considerable flap over what the Federal Government should and should not do, what the Federal role is and isn't. But on this issue, it seems to me there is agreement. Along with helping the disadvantaged, generating and sharing information about the condition and progress of education and about what works is a primary role of the Federal Government. Since 1867 the Federal Government has been charged with this basic responsibility, and no one has ever questioned the appropriateness of it.

This reauthorization is extremely important. It will be effective through fiscal year 1997 or 1998, nearly the entire decade, a decade of enormous challenge. Information about where we stand and what works is absolutely essential to our being able to achieve the national goals. We in the Department don't know now what we might need to do by 1996 to provide that kind of information and help people make the best use of it. The current statutory authority for OERI, when it is accompanied by adequate appropriations, allows us a good deal of flexibility with regard to the activities we can support. We need to be sure that we have this and more flexibility in the future, so that we can act boldly, because that's what it will take from all of us.

We will be sending you our legislative proposal—within the next several weeks. Today I would like to talk about four of the most important aspects of the bill we will transmit.

First, we will include in our reauthorization bill the provisions related to The Nation's Report Card (the National Assessment of Educational Progress) that were included in our AMERICA 2000 Excellence in Education Act. Specifically, we are requesting the authority to:

—Collect State-level data in grades four, eight, and twelve, in reading, writing, math, science, history, and geography on a regular basis, beginning in 1994;
—Provide Federal financial support to the States for those costs in excess of $100,000 associated with their administration of the State assessments;
—Collect assessment data annually; and
—Permit the use of NAEP tests at district and school levels by States that wish to do so.

Measuring the results of education in a way that allows parents, teachers, schools, and communities to compare results and thus insist on change when the results aren't good enough is the only way we can make our schools more accountable, and

it's essential to our achieving the national goals. That is why it is a key aspect of the President's AMERICA 2000 strategy. And for now, The Nation's Report Card is the only vehicle available for providing truly comparable data on what students know and are able to do.

The value of the State-level assessments is already evident. Immediately upon the release of the eighth grade math data last week, States began to call for major review of, or changes in, their math programs. State assessments at all three grades in all five core subjects ought to become a regular occurrence. But think how much more valuable last week's release would have been had it been data on individual districts or even schools. The response of the DC Public Schools to the release, as it was reported in *The Washington Post*, illustrates my point. Because the District of Columbia was able to participate in the 1990 trial State assessment, it was able to receive district data. And immediately officials began to call for improvements. Providence and Kansas City, all school districts, and even schools, should have the same opportunity to learn how they measure up in comparison to others.

We need to have these new NAEP provisions in place by the end of the year, because we must contract very early next year for the 1994 assessments. However, as I alluded to earlier, this NAEP expansion is not enough.

As a Nation, we need to define what it is our students need to learn, what it is they need to know and be able to do to live and work effectively in today's world. We need standards that incorporate both knowledge and skills, to ensure that, when they leave school, young people are prepared for further study and the work force. We also need a nationwide examination system based on the standards, designed to foster good teaching and learning as well as to monitor student progress. As Governor of Tennessee trying to help local school districts design programs for better schools, I found one of the most difficult problems we faced was providing parents with reliable information about how much math or English their children knew. Armed with this information, parents and communities will no longer be able to think, "The Nation is at risk, but I'm OK." Schools will change when families and counties want them to change, and the first step is convincing people that change is needed.

This is why the President's AMERICA 2000 strategy calls for the development of World Class Standards and American Achievement Tests. We are not proposing that this be a Federal effort, that these be Federal standards, or that a Federal test be developed. We are not even proposing a single examination, but rather a system of tests linked in some manner so that they all measure performance against the same standards. I said earlier that we don't know today what we might need to do in 1996. This is an example. Just a few years ago national testing would have been unheard of, but there is a growing consensus today that the country needs this type of voluntary national assessment system. The Federal Government must be able to assist in the national effort to develop these standards and tests, and we need the flexibility to work with the National Education Goals Panel, the National Governors' Association, and others.

This flexibility is the second major aspect of our legislative proposal that I'd like to discuss today. We will propose various changes to enhance the Department's ability to respond with the appropriate information or assistance to the enormous changes we expect to see during the term of this reauthorization. The law currently requires that we spend at least so much each year for research centers, regional laboratories, field-initiated studies, and ERIC clearinghouses. Funding of each of these activities should be determined in the context of all that needs to be done. The current mandated minimums should not be replaced with new ones. The Department must be able to look beyond this set of mechanisms for supporting research and development, providing technical assistance, and sharing information. As with developing the standards and tests, we need the flexibility to explore other means, to support other entities, to carry out all these responsibilities.

Along these same lines, we will propose to expand the types of entities eligible to compete for regional laboratory and research center awards. Currently, only nonprofit organizations may apply to operate regional laboratories, and only institutions of higher education may receive research and development center awards. Removing these restrictions would allow us to involve the expanding cadre of highly competent researchers, who could help improve our schools, but are presently excluded from our major research and development efforts. The New American Schools Development Corporation, which will assist in implementing the AMERICA 2000 strategy by funding the R&D Teams that will develop designs for a New Generation of American Schools, plans to include corporations, universities, think tanks, school innovators, and others on these teams. The Office of Educational Research and Improvement will also be supporting work that will help in creating

"break-the-mold," schools. Already, work supported by OERI has helped us question many out-dated assumptions about schooling and Provided some of the best ideas that point the way to the radical changes AMERICA 2000 aims to help bring about, but OERI should be able to access and support with taxpayer funds the same breadth of expertise the Corporation can reach with the Private sector funds it will raise.

The third issue we will address in our reauthorization proposal is OERI's priorities. The list of priority research needs in the current law is a "laundry list." It covers nearly every conceivable education issue or problem, and the result is no priorities at all. We will propose a new, more coherent set of priorities. It will help us focus on the results of education, on radically improving students' learning, on creating Schools up to the task of educating children for adult life in the next century, and on improving community and family support for children and their education. Because they are designed to help move the country toward the national education goals, we have used as our framework the four tracks of AMERICA 2000:

—Improving today's Schools, making them better and more accountable for results, so that when today's students leave school they are prepared for further study and the work force;
—Inventing a new generation of American schools to meet the demands of a new century, schools that are the best in the world, Schools that enable all their students to reach the national education goals;
—Drawing adults back to School, encouraging all Americans to engage in life-long learning, because education is key both to making a living and to having a more interesting and fulfilling life; and
—Making our communities places where learning can happen, improving the 91 percent of time that children spend outside of school.

Finally, in our reauthorization proposal we will give greater stress to OERI's responsibility to assist in applying the knowledge gained from research and statistical findings. Sharing information and "getting research into practice" has long been a concern. We have begun to explore new ways of helping parents, teachers, principals, business and community leaders, and others use what we learn from research and statistics to make specific changes. For example, we will announce a competition next week for which we will specifically invite proposals for the development of videotapes to provide parents, teachers, and the public with information about strategies that research shows will be effective in helping to achieve the national goals. We want to do more of this.

Before I close, I want to mention the study of the condition of education research that is being done for us by the National Research Council. An interim report is due this fall, and the final report will be available by next spring. A similar study a assumptions about schooling and provided some of the best ideas that point the way to the radical changes AMERICA 2000 aims to help bring about, but OERI should be able to access and support with taxpayer funds the same breadth of expertise the Corporation can reach with the private sector funds it will raise.

The third issue we will address in our reauthorization proposal is OERI's priorities. The list of priority research needs in the current law is a "laundry list." It covers nearly every conceivable education issue or problem, and the result is no priorities at all. We will propose a new, more coherent set of priorities. It will help us focus on the results of education, on radically improving students' learning, on creating schools up to the task of educating children for adult life in the next century, and on improving community and family support for children and their education. Because they are designed to help move the country toward the national education goals, we have used as our framework the four tracks of AMERICA 2000:

—Improving today's schools, making them better and more accountable for results, so that when today's students leave school they are prepared for further study and the work force;
—Inventing a new generation of American schools to meet the demands of a new century, schools that are the best in the world, schools that enable all their students to reach the national education goals;
—Drawing adults back to school, encouraging all Americans to engage in life-long learning, because education is key both to making a living and to having a more interesting and fulfilling life; and
—Making our communities places where learning can happen, improving the 91 percent of time that children spend outside of school.

Finally, in our reauthorization proposal we will give greater stress to OERI's responsibility to assist in applying the knowledge gained from research and statistical

findings. Sharing information and "getting research into practice" has long been a concern. We have begun to explore new ways of helping parents, teachers, principals, business and community leaders, and others use what we learn from research and statistics to make specific changes. For example, we will announce a competition next week for which we will specifically invite proposals for the development of videotapes to provide parents, teachers, and the public with information about strategies that research shows will be effective in helping to achieve the national goals. We want to do more of this.

Before I close, I want to mention the study of the condition of education research that is being done for us by the National Research Council. An interim report is due this fall, and the final report will be available by next spring. A similar study a number of years ago was very helpful in improving the National Center for Education Statistics.

Mr. Chairman, I look forward to working with you on this reauthorization. I will be happy to respond to your questions.

Senator PELL. Thank you very much, Mr. Secretary.

We will limit ourselves to a 10-minute rule, and I would ask my colleagues not to feel any necessity to fill out the 10 minutes, as we have two more panels to follow.

Am I correct in saying that the real difference between our proposals on the test is that the administration is more for testing the school itself and how they are doing, and we are more for focusing on the individual student; would that be a correct statement?

Secretary ALEXANDER. Mr. Chairman, I am afraid I would not agree with that. The administration agrees with you, that it is equally important, perhaps most important, to give parents a way to find out whether their children know what they ought to know about math, English, science, history and geography, and the idea of an American achievement test would be for parents and communities as well as schools to know that information. So we would agree with you about the importance of that.

Senator PELL. That is very similar to the legislation we introduced in the last Congress, I think.

Secretary ALEXANDER. I believe it is, Mr. Chairman.

Senator PELL. Incidentally, I notice that your proposal talks about English, math, science, history and geography. I was wondering who chose those critical areas and why, for example, foreign languages, civics, government, and particularly the arts were dropped.

Secretary ALEXANDER. The answer to the question is those critical areas are the areas that the governors and the President agreed upon as areas of focus in the summit that they held in Charlottesville about a year and a half ago, and they are therefore part of the national education goals. They are also similar to the areas that the Nation's Report Card uses in the assessments that it has been doing.

Certainly no one that I know of thinks that that is all a school ought to teach, and if I were in a community creating a school, I would certainly want art and music and foreign languages and computer skills and many other subjects offered, many of them required. So that it is very possible that in developing a voluntary national examination system that we may want to include other subjects.

Senator PELL. If my memory is correct, we had moved in the legislation maybe it was 15, 20 years ago now, provisions which became law for metric education, and we also had arts included. I

think both of those were dropped in the last 10 years. And my hope would be that perhaps in the subjects that you mention—English, math, history, science and geography—you could include, for example, that some mention of metric education be included. In English, we could maybe add some language or arts.

Secretary ALEXANDER. Those are excellent suggestions, Mr. Chairman, and we have heard those and will take note of them. I will pass that along to the governors who are working on that and others in the administration who are thinking about it.

Senator PELL. They had actually been part of the old ESEA law, and I know my recollection is correct because I authored both of those provisions.

That's all for the moment. Thank you.

Senator Kassebaum.

Senator KASSEBAUM. Thank you, Mr. Chairman.

I have a statement that I'd just like to be made a part of the record. I would to the Secretary that I certainly appreciate his coming up and lending his leadership and stature to this reauthorization because I think, while it may be regarded as more technical to a certain extent, it certainly is very important to the movement of the educational agenda.

I do apologize for being late. I just caught your last comment, and you have probably spoken at greater length about dissemination. I think this has been one of the issues that has frequently been raised, that with the amount of research and so forth that is done, is it really getting out and being utilized in a way that it should?

I understand some in the House believe that, if we could increase dramatically the number of centers, this would help in the dissemination of information. Did you address that at all in your opening comments?

Secretary ALEXANDER. I did not, Senator, and I don't know the answer to that. I would have to be better informed. I'd like to talk that over with members of the department before I form a judgment, but my instinct would be that just creating more centers wouldn't make more difference; that in this day and time, dissemination of good information doesn't depend upon having it cooked up 15 miles away from you. The quality of the information, the power of the ideas that are generated, and the skill with which the information is disseminated might have more to do with it.

Senator KASSEBAUM. Does it normally go to a State's board of education?

Secretary ALEXANDER. Do you mean the information that is developed?

Senator KASSEBAUM. Yes.

Secretary ALEXANDER. Let me ask Dr. Manno, if I may, to respond to that.

Mr. MANNO. The information is disseminated to a variety of agencies. Certainly, State education agencies, SEAs, are part of one of the audiences that we direct our information to, but that's not all. There are local educational agencies, school districts, that are part of our dissemination system.

So our general efforts are directed toward a variety of groups, including schools and teachers in schools. The Secretary did refer in

the course of his remarks to the national diffusion network, NDN. The NDN has been in existence for a number of years, and that entity has been involved in disseminating information down to the school and teacher level.

The reauthorization proposal, as the Secretary said, will include a more focused effort at dissemination, and we propose to add language to the authorization of OERI that would in fact emphasize this dimension of our work a lot more than it is in the authorization that we presently have.

Senator KASSEBAUM. That sounds very sensible, I think. Perhaps it is a more direct focus that would be useful, because sometimes we all get so much information that it is hard to sort through the priority items or give attention to any one part of it at a given time. So that sounds very useful to me.

Just a moment on the test. In a national test, so to speak, will there be any differences that will be allowed for a district or even by a State? Has there been any thought given to that?

Secretary ALEXANDER. Yes, there is a good deal of discussion about that, Senator, and there are different opinions about that. And I think that is healthy and good, and we ought to talk those through over the next several months.

The way we're thinking about it in the administration now is that it is a voluntary examination system, so that in Kansas or Wichita, parents would say we'd like to know what our children know about math and science, and then they would decide what sort of report card they wanted to use, and they might choose to use this national examination or they might choose to use another one. What we would hope the governor would do or the Senator would do is at least provoke the debate and say, well, if you use another sort of Kansas test, at least make sure it's to the same new, higher standards that the national achievement test is.

We in the administration think that that will cause communities and States to buy into the examination process, pay more attention to it, take it more seriously, that that is the way America works.

Now, there is a contrary view which says that it would be better to simply go ahead and not only agree on national standards but agree on a national examination that is given to everyone and let them find out where they stand. Those are competing points of view, and as I understand Senator Pell's proposals in the past, he has argued toward a national examination that everyone would take. Other countries in the world do that, and many people feel that this country should.

Senator PELL. The reason why I had it voluntary in the original legislation was that that would act as an engine for the school committees, which I think usually have too much to say in these things, who they would want their children to be able to do well on the test in competition with other schools.

I am much more supportive of the way it is moving along with some compulsiveness in it.

Sorry to interrupt, Senator Kassebaum.

Senator KASSEBAUM. Not at all. I think it is helpful to have that point of view.

So it would be your thought that perhaps analyzing of the tests and putting together the results would be done within the National Center for Education Statistics?

Mr. MANNO. That's correct.

Senator KASSEBAUM. So that work would all be done there, the calibration of the tests and so forth?

Secretary ALEXANDER. Well, it could be done there. All that is part of the discussion that we need to talk about. Let's take mathematics, for example. The mathematics teachers have been working for a few years to develop the curriculum standards that might be used in a national examination of mathematics and are now developing assessment strategies.

I suppose it is possible that commercial publishers or someone outside the government might develop an examination based on those standards and those strategies that many educators and many States and many communities would find to be an excellent national examination based upon new national standards. There might be more than one such examination, and there might develop a consensus that all of us were to this standard, in somewhat the same way that you are able to compare college admission tests, which are different types of tests—the ACT is one, the SAT is another—and college administrators can compare scores from one to the other. So there might be, and we think there probably should be, more than one option for a State or a community to consider.

Mr. MANNO. Even with reference to the way the Nation's Report Card, or NAEP, is presently conducted, it is not conducted in a way which is exclusively an activity of the Federal Government. For example, on the development of consensus on what it is the test objectives should be, that is in fact an initiative that has been carried out through the Council of Chief State School Officers.

So even at present this activity is not something which is the exclusive preserve of the Federal Government, although in fact when it comes to funding, most of the funding does come from the Federal Government. So there are ways in fact in which the present test is conducted that involve a variety of outside groups in the development of consensus.

Senator KASSEBAUM. Actually, that has been done since 1969; isn't that right?

Mr. MANNO. That's correct.

Senator KASSEBAUM. So I would think that there would be much from that effort that would really have a significant input into just how the further testing should evolve, because NAEP has been pretty much what we would like to see, only broadened; is that correct?

Secretary ALEXANDER. Many people feel of that type of test, the Nation's Report Card or the NAEP tests are the best tests. I know as a governor, when I was looking for some reliable information about what was going on that I could rely on, I had confidence in that. Now, there might have been other tests in which I could have had confidence, but I had great confidence in that. So we have had 20 years of experience with it, and obviously the experience of NAEP, the Nation's Report Card, should be a major contributor to any effort we make to produce a so-called American achievement test or a national voluntary system of examination. Now, just how

to do that is something we need to talk through, and we need to talk through it publicly and openly so lots of people are involved, because lots of people have opinions about it.

Mr. MANNO. It is important to keep in mind here, too, that NAEP as it presently exists only presents us with data that suggests trends, national trends and most recently, although we really don't have a trend line yet on this, State data.

What the administration is interested in discussing and engaging in a discussion with involves the development of these American achievement tests that would begin to give us data on a student level, on a district level, and on a school level.

So what we would hope to continue to do would be to preserve the integrity of NAEP as it presently exists when it comes to maintaining trends over time, but begin to work at developing a system that would provide us with a lot more information on a very specific level that would be useful to parents and to teachers and to superintendents.

Senator KASSEBAUM. Thank you very much. I appreciate it.

Senator PELL. Thank you very much, Senator Kassebaum, and your statement will be included in the record.

[The prepared statement of Senator Kassebaum follows:]

PREPARED STATEMENT OF SENATOR NANCY LANDON KASSEBAUM.

It is a pleasure to welcome Secretary Alexander once again to the subcommittee, as we consider the OERI reauthorization. The collection of data and support of research are functions which the federal government is uniquely suited to do. Whatever disagreements may arise about the appropriate role of the federal government in education, I think all would agree on this point.

National level statistics and research provides a valuable tool to educators who want to apply successful practices from other areas, to all who are interested in how they stand relative to others, and to Congress as we try to gauge the impact of federal legislative proposals.

As with any tool, it is effective only to the extent it is generally available and properly used. The wealth of information generated by the federal government must be in a form which is understandable and put in the hands of those who can apply it. An important focus of this reauthorization process will be assuring that data and research will not be left on the shelf to gather dust but will rather serve as a stimulus to further reform.

This reauthorization will also provide an opportunity to consider questions related to testing at the national level, including the role of NAEP (National Assessment of Educational Progress) as well as numerous new proposals for national testing. Like any other form of data, test results are only useful if those actually in the classroom understand them and find them relevant.

I look forward to hearing the perspectives of today's witnesses on these and other issues related to OERI.

Senator PELL. Senator Bingaman.

Senator BINGAMAN. Let me ask a few questions about your National Center for Education Statistics. Do we have a head for that?

The head is designated the commissioner of education statistics. Is there a person in that job?

Mr. MANNO. Yes, there is an acting commissioner in that job, and I have the responsibility of recommending to the President a permanent commissioner who would then make a recommendation to the Senate for that.

I am about that, but I haven't made that recommendation yet.

Senator BINGAMAN. How long have we had an acting person in that position?

Mr. MANNO. We have had that for somewhere between two to 3 years. It was a result of the last piece of legislation that the Congress passed on the center; it was part of the Hawkins-Stafford Act of a couple years ago.

Senator BINGAMAN. We left it vacant for two or 3 years because of that legislation?

Mr. MANNO. Well, the legislation actually specified that what the department was to do was appoint an acting commissioner, so we followed through with the full intent of the legislation; and the legislation further specified that as of June of this year, late June of this year, the Secretary is to forward, as he said, to the President a formal recommendation about who the next—first, actually—commissioner of the center would be. So we are in full compliance with the law.

Secretary ALEXANDER. Emerson Elliott is the acting commissioner, and a distinguished career person and, most people believe, doing a very good job in that respect.

Senator BINGAMAN. Let me ask about the budget. I have been led to believe that a very major problem with the National Center for Education Statistics is that in the budget that is submitted by the administration each year and approved by the Congress, we do not have a line item for salaries and expenses in that office, and accordingly the head of it is not able to hire people and winds up having to farm everything out, and that has severely impeded the ability of that office to develop a professional staff.

Am I wrong or right?

Mr. MANNO. You are correct in saying that the budgets that have been submitted do not include a specific S and E line item for the national center. I don't think one can necessarily then go on to conclude from that what you have concluded. I think it has been the position of the department that dealing with the national center in the way that it is dealt with—which is to say an entity that is part of OERI—actually allows the center to be more flexible when it comes to the use of general department S and E funds and general department hiring authority with reference to FTEs.

You are accurate, though, in saying that the center has had to rely upon the work of outside contractors as well as interagency agreements with Bureau of Census and others when it comes to the collection of statistical information.

Senator BINGAMAN. So you don't think that the lack of this salaries and expenses line item is an impediment to the effectiveness of that office?

Mr. MANNO. Not to the ultimate effectiveness of it.

Senator BINGAMAN. Well, I have been advised that there are three people there now focused on long-term planning and that

they are all one-year hires or fellows of one kind or another; that there is nobody who is sort of a full-time civil servant for the duration, looking at long-term issues in that office.

Am I wrong?

Mr. MANNO. I would beg to differ on that issue. There is in fact an associate commissioner who is appointed at the SES level, whose major responsibility includes—the specific title is not long range planning, but in fact that's what she does—long range strategic planning.

Senator BINGAMAN. How much staff does she have?

Mr. MANNO. I'd probably have to check to be exact, but I do know that she has at least two people who work under her. Now, there is in the center a one-year arrangement for certain individuals who come into the center for 1 year under the general auspices of the fellows program, and it is also true that some of those fellows assist her in her work. But the specific issue that you raised with reference to strategic planning does have an associate commissioner concerned with that issue.

Senator BINGAMAN. Well, I guess my point and what I'm driving at with this set of questions is the ambitious plans, which I heartily endorse, sound to me like they are going to require a dramatic increase in funding and staffing—at least staffing—in that office.

Do you agree with that?

Mr. MANNO. Yes, I think that is accurate.

Senator BINGAMAN. OK. Also with NAEP, last year NAEP did a State by State comparison for 4th grade math, as I understand it. You are now suggesting that in 1994, they will do a statewide State by State comparison—is that what I understand—in the five subjects, at three levels?

Mr. MANNO. The most recent release of data in math, which occurred just about a week ago, was for 8th grade math.

Senator BINGAMAN. Oh, 8th grade. Excuse me.

Mr. MANNO. Eighth grade. The plan in 1992 is to have State data in math in the 4th and 8th grades and reading on the 4th grade level.

Senator BINGAMAN. Right, but 1994 was what I was asking about. I thought, Mr. Secretary, you had said that you wanted the five subjects—math and science and English and history and geography—tested at 4th, 8th and 12th grades on a State by State basis.

Mr. MANNO. That's correct.

Senator BINGAMAN. Do you have an estimate as to what that will cost? What kind of an increase in NAEP's budget will be required to accomplish that?

Mr. MANNO. The center staff do have estimates of what that would cost. I don't have that information with me, but I'd be pleased to submit it for the record.

Senator BINGAMAN. I'd be glad if you would because my sense is it will be a dramatic increase——

Mr. MANNO. You are correct in saying that.

Senator BINGAMAN [continuing]. And I think that's fine. I'm not arguing with that, but I just want to be sure we are all in agreement on the amount of money that the administration is going to have to request this fall or in the budget that is prepared this fall

and submitted in January if you are going to follow through on that. What is the NAEP budget today, roughly, do you know?

Mr. MANNO. I don't know off the top of my head. I have a table here, and I can check it for you. [Pause.] It is $28 million.

Senator BINGAMAN. So you'll have to add $100 million or so to that this next year, won't you, if you are seriously going to pursue the goal that you have described?

Mr. MANNO. It certainly will involve a substantial increase. Whether it is $100 million or not, I hesitate to say, and as I said, we'd be pleased to have that figure put into the record.

The fiscal year 1992 request for NAEP is just a little over $28 million.

Senator BINGAMAN. OK. Now, let me be clear in my own mind as to the distinction between assessment, which is what you are talking about NAEP doing, to continue to assess trends, and the development and implementation of a national testing system which will include, or at least to my mind I think it would involve even a much greater level of effort——

Mr. MANNO. That's correct.

Senator BINGAMAN [continuing]. Than the national assessment that NAEP is now doing.

Mr. MANNO. That's correct.

Senator BINGAMAN. Who is going to fund that, and where is the money coming from for that?

Secretary ALEXANDER. Senator, if I may, I don't think we know the answer to that yet. I think that is part of the discussions that we should be having over the next 6 months in this interim council on standards and testing is taking each subject, math for example, and saying what would be the best way to develop a national examination system in mathematics, and who should do it, and who should pay for it.

There may need to be a Federal contribution. There could be State by State contributions. There could be interstate compacts by States that sought to do this. And then there is a commercial publishing world, and then there are foundations that might seek to be involved, and then there are the users of the examinations, ultimately, who would pay the costs of the administration, which in the end get to be among the largest costs.

So the answer is if every child in the country in the 4th, 8th and 12th grade took an American achievement test in those five subjects, the total cost would be a huge cost, but it might or might not be a total cost to the Federal Government.

We are prepared to recommend that the Federal Government pay for whatever it ought to pay for, but we want to be careful about what that should be.

Senator BINGAMAN. Your reference earlier to $100,000 per State—what is that for?

Mr. MANNO. That refers to the costs associated with State administration of a State assessment. Our proposal which is up before the full committee with reference to NAEP involves the Federal Government picking up all of the costs associated with State assessment except for the first $100,000. The $100,000 figure actually is a carryover from the most recent experience that we have had on the math assessment.

Senator BINGAMAN. So you are suggesting that the States would pick up $100,000 of expense, and then the Federal Government will pay the rest.

Mr. MANNO. But this is for the assessment and not for anything related to the test.

Mr. MANNO. That's correct. It is costs associated with administering these tests in the States.

Senator BINGAMAN. My time has expired. Thank you very much.

Senator PELL. Thank you very much, Senator Bingaman.

Senator Jeffords.

Senator JEFFORDS. Thank you, Mr. Chairman, and I have a statement I'd like placed in the record.

Senator PELL. Without objection, it will be inserted in the record, along with a statement from Senator Coats and any other members who so desire.

[The prepared statements of Senators Jeffords, Coats, and Hatch follow:]

PREPARED STATEMENT OF SENATOR JEFFORDS

Mr. Chairman, thank you. Quickly, let me thank our witnesses for coming before us today. I look forward to their testimony.

Reauthorization of the Office of Educational Research and Improvement (OERI) may not get headline news coverage. Yet, more times than not, the results of their work do reach the media. OERI provides the necessary studies that spur this Congress and the nation to implement the reforms. So I consider this reauthorization of great importance for that very reason.

OERI's charter, to provide leadership in the conduct and support of scientific inquiry into the educational process must be firmly restated during this year's reauthorization.

The task of education has become a greater challenge because of increasing numbers of children from economically and socially disadvantaged homes. The knowledge base for this new challenge is often inadequate.

Policymakers need a source of objective and trusted information about education. The neutrality of the data systems provided under this Act from the National Center for Education Statistics, to research centers in universities, to the regional educational laboratories and the grassroots research needs of educators and policymakers as well as the accessibility of the ERIC system are critical components of the system.

Congress depends on the results of these and other studies when fashioning new policies or reevaluating old ones. I, therefore, look forward to working with the Committee and the Administration to reauthorize a strong, critically needed research arm for education information and strategy.

PREPARED STATEMENT OF SENATOR COATS

Mr. Chairman. I am pleased to welcome Secretary Alexander to the committee for the second time this week. His appearance is evidence of the strong commitment of the administration to work with the Congress to further our Nation's educational agenda.

I think it is quite appropriate that the hearing on the Office of Educational Research and Improvement follow a discussion of the Federal proposals to reform and restructure our educational system. The President and the governors set ambitious goals which have galvanized the debate and fostered new resolve to tackle the perplexing issue of mediocrity in our system. Now more than ever we need to focus on the basic questions of what works in the classroom and why.

I believe that OERI can play an important role not only in assessing educational progress but in identifying and disseminating information on successful learning, teaching and structural reform. It seems appropriate to me that OERI play a role in helping communities to apply these principles in launching reforms at the local level. Work by OERI will also be an important component in determining how we meaningfully evaluate our success in reaching our common goals.

I appreciate the opportunity to hear from Secretary Alexander and our witnesses today.

PREPARED STATEMENT OF SENATOR HATCH

I would like to again welcome the Secretary here this morning. He is becoming a very familiar face here at the committee. I think that is wonderful. It shows how committed he and the President are to education and how willing he and the President are to work with members of Congress.

The Office of Educational Research and Improvement serves a vital role in the Department of Education. It collects, analyzes, and disseminates data needed for those who assess progress as well as those who must make informed decisions regarding educational policies and practices. OERI is not just for the university scholar; it is also for the state superintendent and local school board member.

The research performed under the direction of OERI is helpful to teachers and school leaders throughout this country to improve teaching and learning in our schools. Only by knowing what works and what doesn't can we make appropriate changes to ensure student progress and achievement.

The National Assessment of Educational Progress, a key function of OERI, helps us to evaluate our progress as a nation in the subjects of reading, mathematics, science, writing, history, and geography. One major topic that will be addressed in this reauthorization is the development and implementation of national tests. I intend to follow this discussion closely. We need to look carefully at the implications and the benefits of such a system of tests. We should proceed cautiously and listen to all points of view before making a decision.

The functions of OERI are very appropriately assigned to the federal government. Data collection which is uniform and comparable can only be attained by establishing national definitions. Research on education should be coordinated and disseminated from a national perspective, so that valuable research funds are not used to replicate research which has already occurred elsewhere in our nation.

I trust that we can work together on this reauthorization to ensure that the Department of Education has the flexibility necessary to provide states and local leaders, as well as national leaders, with the information, the ideas, the research, the data, and the methods of dissemination necessary for our schools to properly meet the needs of all our children.

Senator JEFFORDS. I'd like to focus on a different aspect. I am excited about looking forward to the future and breaking out of the mold of our present school systems and seeing what opportunities are available to us to make some leaps forward in our educational system.

Over the years I have had an opportunity through various interests to take a look at some of the work that has been done with computers and software. The Plato System in particular I got interested in when I was working with the CETA program, Job Training Partnership programs, and I was deeply impressed at the incredible advancement that was allowed some people who were alienated from the school system.

Subsequent to that, I became interested in the "Writing to Read" program at IBM and followed that to some extent and became quite friendly with Dr. Martin, who was the innovator there. I suggested to him, for instance, that they ought to try something in a foreign country to see whether we could help and assist Third World countries. IBM created a program with the Zulus in South Africa and were amazed that their children could learn how to speak and read English very quickly. He was invited over, and he had some trepidation, but he was made an honorary chief.

I have talked with the people who are with the "Writing to Read" program and with John Akers who is in my office. I asked him this question, to which he did not have an answer. We take our young people, we train them in pre-kindergarten to be able to read and write, and we dump them onto some poor first grade teacher who is used to teaching the alphabet to start with; what happens to them, and are there any programs that are following up on these children, to see whether we have any way to take advantage of this opportunity that we gave these young people in kindergarten? Is there any follow-up? John Akers was unable to answer that.

I wonder, is there anything going on in your office, Dr. Manno, in this regard to take a look at what kinds of opportunities these leaps forward have presented under certain circumstances and to see whether or not you can develop total curricula based upon the software. I don't know whether there is anything that is beyond the "Writing to Read" program in the software area that is being utilized or anybody looking at designing programs with teachers, obviously, as assistants and aides. Can we design something which can give us the kinds of leaps forward we got with Writing to Read?

Mr. MANNO. Yes, we are. There are a couple of different ways in which we are doing that in OERI. One way involves supportive work in one of our research and development centers, the center on technology, which is at Bank Street in New York. They have been involved for a number of years, working with a variety of groups, looking at issues like that.

We could provide you directly or for the record a detailed description of the work they have supported and the kinds of things they have done.

Another way goes back to an earlier issue that we were talking about, and it involves the work of the National Diffusion Network, whose main function is to collect information about programs that work and then disseminate those programs, get those programs out to individuals in schools, individuals in State education agencies and so on.

There is in the "America 2000" plan—and perhaps the Secretary would like to talk a little bit about this—an activity that we are describing as bringing America on line that relates very much to what you have talked about.

Secretary ALEXANDER. Senator, yesterday I was in Public School 25 in the South Bronx for a little visit. It is a school that is 96 percent Hispanic, and almost none of the parents of the children in that school speak English very well, so they have some considerable challenges and really some advantages because all the children come out of there speaking two languages, one of them English, which is a significant advantage.

In visiting a kindergarten there and also in the first grade, there was the "Writing to Read" program there. The principal was telling me they were using the "Writing to Read" program to improve the learning of native languages, and because they were working hard on mathematics, they were using the "Writing to Read" program to improve the English language facility of those students in terms of mathematics, so they could solve mathematics problems and improve their test results.

Now, it seems to me our job, I guess, here with the OERI is to let other schools whose students are similarly situated know how they in P.S. 25 adapted the "Writing to Read" program to help children solve mathematics problems better.

Senator JEFFORDS. My concern is that the computer industry, which is not having the most profitable times, is not likely to spend large sums of money to develop unprofitable items. Whether or not these kinds of things can be profitable in terms of the computer industry concerns me, and as to whether or not there is sufficient work being done through your office and others to better understand the potential and bring that technology forward. We need to figure out just how we can do that and what the costs would be.

Secretary ALEXANDER. Well, that goes to the American on line proposal, which is a little noticed provision, really, in the "America 2000" proposal, but potentially one of the most exciting.

I think all of us feel this need that you have described. The Librarian of Congress has talked to me about the vast amount of information that he has there which ought to be more useful to classroom teachers and to principals and to schools, or to individuals who want to learn more.

So the idea of American on line would be to create a sort of quasi-utility that would find a way to gather lots of information that could be transmitted visually to classrooms around America and that would help them learn about activities such as the one that you describe, to help them pull that down to their individual classrooms.

The example we often use is the National Geographic Society, which has most of what it has done on a couple of interactive videodiscs. Now, that's a little bit of overstatement, but not much. And geography is one of our national goals, yet most of our 110,000 schools are not equipped to include within their curricula or what they are doing the 100 years of National Geographic films or prints that could be an essential part, or the "Writing to Read" program or what is learned from that. In America on line, we have recommended in our appropriations $5 million to begin to develop and plan for this utility and make ideas such as the ones you have suggested more known.

Senator JEFFORDS. That's encouraging to hear; I'd like to have more information on that because the information systems, fiberoptics and so on, give us great potential to be able to make these kinds of things cost-effective.

Secretary ALEXANDER. Absolutely.

Senator JEFFORDS. I know because I worked out a program with Plato for my daughter, but the phone bill was so high that we had to terminate it. Of course, I found out it was probably high because she was playing the games that were available on the system probably more than the course. Anyway, it led me to understand how important it is that we develop these informational systems and networks, satellites or fiberoptics or whatever, to make this an affordable option. So I'm pleased to know that some work is going on.

Mr. MANNO. Actually, you'll have appearing before you shortly an executive director from one of our regional educational laboratories who has been involved in this work, and she surely would be able to give you some other examples about how the support money that goes to the regional educational laboratory through OERI has led them to begin to get involved in these kinds of activities.

Senator JEFFORDS. Vermont is moving ahead fairly quickly in these kinds of things. I have been on some of the interactive television programs, working with New England Telephone. It is an incredible opportunity.

Mr. MANNO. We also have within OERI, for example, the Star Schools program. We are charged with administering that program. So there are a variety of activities that we support that get at this issue from a variety of different perspectives.

Senator JEFFORDS. My last question is has any study been done following up on what happens to the "Writing to Read" students?

Mr. MANNO. I don't know that we have supported any follow-up studies to that. I could certainly check on that.

Senator JEFFORDS. I'd appreciate knowing because it seems to me it is incredibly important to find out. It would be interesting to see what happens when you drop these kids into a conventional school system with those advantages, and whether it is all gone after a year, whether there is any effort to see whether similar kinds of activities can continue that leap forward that they got——

Mr. MANNO. I do know that there has been research done just on that specific issues, but whether we have supported it or not, I don't know.

Senator JEFFORDS. I appreciate that.

Thank you very much, Mr. Chairman.

Senator PELL. Thank you very much.

The record will stay open for questions from any of the members. I'd like to move on to the next panel.

Thank you very much indeed, Secretary Alexander, for being with us, and Dr. Manno. We look forward to working with you in the committee.

Our next witnesses are Dr. Jeri Nowakowski, executive director of the North Central Regional Educational Laboratory, Oak Brook, IL, on behalf of the Council for Educational Development and Research; Dr. Arthur E. Wise, president, National Council for Accreditation of Teacher Education, on behalf of the American Educational Research Association; and Mr. Nathaniel M. Semple, vice president and secretary, Committee for Economic Development.

We'll proceed with Dr. Nowakowski first—and I think you know the ground rules. You are limited to 5 minutes, and the Senators are limited to 10 minutes for questions, and we may have a second round if necessary.

Dr. Nowakowski.

STATEMENTS OF JERI NOWAKOWSKI, EXECUTIVE DIRECTOR, NORTH CENTRAL REGIONAL EDUCATIONAL LABORATORY, OAK BROOK, IL, ON BEHALF OF COUNCIL FOR EDUCATIONAL DEVELOPMENT AND RESEARCH; ARTHUR E. WISE, PRESIDENT, NATIONAL COUNCIL FOR ACCREDITATION OF TEACHER EDUCATION, WASHINGTON, DC, ON BEHALF OF AMERICAN EDUCATIONAL RESEARCH ASSOCIATION, AND NATHANIEL M. SEMPLE, VICE PRESIDENT AND SECRETARY, COMMITTEE FOR ECONOMIC DEVELOPMENT, WASHINGTON, DC

Ms. NOWAKOWSKI. Thank you, sir.

Good morning, I am Jeri Nowakowski. I am the vice chair of the Council for Educational Development and Research and the director of the North Central Regional Lab outside Chicago.

Our lab serves seven Midwestern States that have 4,300 school districts, 30,000 school buildings and about 20 percent of the Nation's students.

The council is composed of the 10 educational laboratories and a number of the Nation's largest university-based research centers.

Mr. Chairman, let me concern a little about what labs do. Labs get the latest research results to educators. We do so through print, through programs, and in fact last year in conjunction with PBS, our laboratory sent out a nine-part television series on research that can be used to help restructure schools. But we do more than just deliver information. We tailor it for the use of educators. We take it into schools, and we empower teachers and administrators to use it. I'll give you an example of that in a moment.

The authorizing statute that created OERI also authorized laboratories and research centers, and because of this we have given serious thought to what we should recommend to this committee. Many people are going to come to you with different solutions for America's educational problems.

Let me begin by stating what I don't think the problem is. The problem is not that we do not know what is excellent in education, nor that we lack for examples of excellence in education. In fact

this Nation has some very excellent schools, there are excellent teachers, and there are high-achieving students across the continent. We do have proven programs and strategies, and many of them have been produced and disseminated as a result of your Federal investment.

But the problem is that those solutions and that excellence are not in every school in America.

Second, the problem is not deciding how to create or organize or reorganize the Federal R and D system. You do have an infrastructure; it does work, but it needs to be a good deal stronger. It needs the capacity to penetrate deeper and broader so that all of the 83,000 public schools and the 5,000 private schools in America can benefit from it.

The existing Federal research and development system is already developing and delivering strategies that work. Let me give you an example. In Wisconsin, isolated, small, rural schools find it next to impossible to keep teachers abreast of the new research on reading and to update their skills. Our laboratory used the current research on reading, much of it from the research center at the University of Illinois, to produce a reading program and a teacher training component that used distance technology. We used narrow-band radio and TV, an electronic network, telephone conferencing, VCRs and computer classrooms, and we delivered that program to 19 rural Wisconsin schools.

Did we increase 3rd graders higher order reasoning skills in reading? Yes, we did, and we did so in 3 years. Did we change the way teachers taught? Yes, we did, and in fact independent evaluators—that is, the NDN and the program effectiveness panel—just validated our program as a program that works.

Our real challenge in the Midwest is getting research findings and programs that work into all 30,000 of our schools. The way to meet this challenge is to increase our focus on the "D" in "R and D". Real improvement in American education occurs classroom by classroom and school by school, and the Federal R and D system needs to develop strategies to speed up that change and to get proven, break-the-mold programs into greater numbers of schools.

Consequently, sir, we respectfully make the following recommendations. First, that Congress define the research agenda and demand stability in solving the problems that it identifies; second, that the current research centers should be reduced in number and increased in size and capacity so that they can put a critical mass to focus on these problems; third, that for some of our critical national problems, Congress consider creating "super institutes", for example, to deal with educating disadvantaged children; fourth, that regional educational laboratories be maintained and that they be given some stability in order to keep and attract the quality staff we need to serve schools.

We ask that laboratories be linked through collaborative work with other parts of the educational R and D infrastructure, with the State delivery systems, with intermediate service agencies, and with teacher centers. These agencies form the launching pad for educational extension agencies which provide direct services to schools.

We would ask that you strengthen the authorization for field-initiated research. Don't just limit your investment to regional educational laboratories and research centers or institutes. Provide provisions for field-initiated research to promote work by minorities, by researchers who have been in the field for less than 5 years, for teachers and buildings.

Most important, we urge that you concentrate Federal resources on a critical mass of effort to effect large-scale educational change and to deliver it.

Mr. Chairman, we will do everything we can to aid you and the committee in forging a Federal research and development system that can truly reform American schools.

Thank you.

Senator PELL. Thank you very much indeed. Dr. Nowakowski.

[The publication entitled "What We Know About Mathematics Teaching and Learning," by Nancy Kober, submitted by Ms. Nowakowski is retained in the files of the committee.]

[The prepared statement of Ms. Nowakowski follows:]

PREPARED STATEMENT OF MS. NOWAKOWSKI

Good morning. I am Jeri Nowakowski, Vice Chair of the Council for Educational Development and Research. I am also Executive Director of the North Central Regional Educational Laboratory, located in Oak Brook, Illinois, outside of Chicago. On behalf of the Council, we appreciate your invitation to appear here today.

The Council's mission is to support the Congressionally created educational research and development institutions. These institutions are working to find ways—either through their own investigation or the evaluation and use of other research—to enable every American school child to succeed.

The Council is composed of all the regional educational laboratories and a number of the nation's university-based educational research centers. Mr. Chairman, let me explain exactly what we do.

The federal government supports a network of 10 regional educational laboratories. These laboratories serve the educational research and development needs of the nation's schools. Each laboratory is governed by an independent board made up of representatives from a fuR spectrum of regional educational interests, including business, all levels and sectors of education, and parents. This regional board is important because it protects us from local and state politics that can bog down long-term change and helps us maintain an independent, objective R&D program.

The reason we are called "laboratories" is that we help schools and educators take the high risks associated with change. Laboratory staff go to a school, school district, or state only when invited, only when the solution to a problem has wide applicability and only when the problem has been defined as a high priority one in the region.

But laboratories do more than give the latest research results to educators. They actually shape those results in such a way that they are useful in solving real-life school problems. Much of the research we use is from the university-based federal educational research centers.

Regional educational laboratories and university-based R&D research centers are two of the six authorized programs in the statutory authority that created the Office of Educational Research and Improvement within the U.S. Department of Education. The other four programs are: (1) the Educational Resources Information Center (ERIC), which serves as the federal archival and retrieval system for educational research; (2) a field-initiated research program that funds meritorious, unsolicited proposals from researchers; (3) the National Assessment of Educational Progress, which is the federal government's assessment of what students know in selected core academic subject areas; and (4) the National Center for Education Statistics, the national educational statistical agency.

The present authorization also permits funding some programs at the discretion of the Secretary of Education pending appropriations from Congress for such purposes.

The authorizing statute that created the Office of Educational Research and Improvement five years ago also authorized the regional educational laboratories and the educational research centers. Because of this, we have given serious thought to what we should recommend to this Committee. Time doesn't permit me to mention anything but highlights ¶of our analysis. However, let me just say that we look forward to continuing discussions with you about the reauthorization of the federal development and research statute.

I begin by stating what the problem is not—all rhetoric to the contrary. The problem is not a lack of excellence in American schools.

This nation has some very excellent public schools. There are excellent teachers. There is excellent support from communities. There are high achieving students. We have programs and strategies that work—most of them produced as the result of your federal investment.

The problem is that our best programs and strategies have not reached all the schools they need to reach. We don't have excellent teachers in every school. Every community is not a nurturing environment for education.

Neither is the problem simply a matter of deciding how to organize the federal educational research and development infrastructure. The present infrastructure works. Yes, it needs to be stronger. Yes, it needs to penetrate deeper and broader so that all of the 83,000 public schools and 5,000 private schools in America can benefit from it. But the Committee does not have to reinvent the system. The existing federal research and development system is developing and delivering strategies that work.

Let me assure you that your investment in the regional educational laboratories is paying off. Where we have had a concentrated mass of talent, knowledge, and risk-taking grit aimed at our most critical education problems, we have had success. Let me illustrate.

In Wisconsin, the laboratory, in cooperation with the state department of education and the educational broadcasting system, set out to attack the chronically low reading achievement in poor, rural schools.

Wisconsin, like other states in my laboratory region, has some schools with so few resources that it is close to impossible to keep teachers abreast of new developments and update their skills. Many schools simply cannot afford to pay substitute teachers while regular teachers attend staff development programs elsewhere.

We used the most current research on reading—much of it from the federal research center on reading—to not only produce a reading program, but also an intensive teacher training component that uses multimedia distance learning techniques.

Did we increase third graders' higher-order reasoning skills in reading? You bet we did. And we did it in three years. Did we change the way teachers taught? You bet we did. Independent evaluators reported significant changes in teaching.

Allow me to use this one program to point out some essential truths about the federal research and development system.

1. Positive school change requires that schools and professionals who work in them have access to sustained technical assistance. We provided the schools using this program with such technical assistance by building a school team. The team consisted of the teacher, the principal and the librarian" resource/ media person. The team worked together and with other teachers. They communicated by computer bulletin boards and telephone. ¶ Teachers did not have to go away to be trained. They were trained while they were teaching. This had the added advantage of allowing them to immediately try a new teaching strategy in the classroom.

2. The real challenge is getting research findings—in usable form—into all of our nation's schools. The way to meet this challenge is to increase focus on the "D" in "R&D." Real improvement in American education will occur classroom by classroom, school by school. The federal research and development system needs to develop strategies to speed up change in a greater number of schools.

Initially, 19 schools participated in our pilot reading program. Now we are planning to disseminate it to the other seven states in our region. And urban schools are demanding it too, so we are adapting it for that environment.

Ladies and gentlemen, this is a "break the mold program." The federal dollars you have provided to the regional educational laboratories are producing other programs just like this. I can give the committee other superb examples if you wish.

3. Developing programs based on research findings requires design teams in the federal R&D system. This country doesn't have a chance of reforming schools if it depends only on the research function. Publishing in prestigious journals does not bring about school reform. It's a necessary step, I agree, but, alone, not enough.

The North Central Regional Educational Laboratory is staffed with a team of talented professionals. Their bottom line is not how much they publish. Rather, their

standards for productivity are: (1) the capability to evaluate research and judge how to use it in schools; (2) the capability to work in teams with a variety of people having diverse knowledge; (3) persistence in adjusting, adapting, and refining a process until it is successful; (4) understanding of the situation in which educators find themselves; and (5) respect, and admiration, even, for the people in the classroom.

But you can't put a team like this together and expect them to deliver on "development" without giving them some assurance that the institution in which they work is stable. The federal research and development effort has been shaken to its very roots by changing political ideologies. Whatever you do, please do not introduce more instability into the system.

We believe that for the strategies, techniques, and programs that form the nuts and bolts of school reform to penetrate deeply into our educational system, there must be: (1) stable institutions so that good people will work in them, (2) barriers to protect research and development from constantly changing short-term objectives, but enough flexibility to respond to fresh leadership, and (3) the definition, by Congress, of generic education problems where the federal dollar could leverage change.

Consequently, we respectfully make the following recommendations:

a. Congress should define the research agenda and demand stability of effort in solving the problems it identifies. b. Current research centers should be reduced in number and transformed into institutes. Institutes should focus on persistent problems such as testing and assessment, professional development for educators, and curriculum issues. But these are only examples. The real issues should be defined by Congress.

Such institutes should be funded at no less than $5 million a year and given at least a 10 year period in which to do their work. The current configuration of 25 tiny centers—many of which are no more than coalitions of part-time researchers working on their pet research agendas—spread across 60 universities, researching discrete topics, and guaranteed only five years of funding cannot produce a body of useful concentrated research.

c. On critical national problems. we urge Congress to consider creating a super institute. We suggest that the super institute work on the problems of educating disadvantaged children. But, we offer some cautions. This institute must have an adequate full-time staff and be linked with other parts of the federal educational R&D system. When examining the costs of such an institute, Congress should weigh them against the benefits as well as the probability of adequate appropriations.

d. The regional educational laboratories should be maintained. However. legislation should more explicitly define their role. The laboratories need more assurance of stability in order to attract quality staff.

The present laboratory regions also need to be maintained. Regionality permits greater sharing and efficiency. Moreover, laboratory governing boards depend on broad regional representation, first to grant them the power to gear Congressional direction to specific regional problems and, second, to create protective barriers from short-term, politically driven goals that can derail a laboratory's work.

The laboratories should be linked, through collaborative work with other parts of the educational R&D infrastructure and with state delivery systems, such as intermediate service agencies and teachers centers. These agencies form the launching pad for education "extension agents," who provide direct services to schools. Every state has such an infrastructure. We should use it.

e. We also urge you to strengthen the authorization for field-initiated research. Please do not limit your investment to regional educational laboratories, research centers or institutes. We suggest that special provisions be made in field-initiated research to promote more work by minorities, researchers who have been in the field for less than five years, and small research teams. This work would be in addition to that done by the general research community.

f. Concentrate federal resources on a critical mass of effort to affect large-scale educational change. We would encourage strategic use of federal resources in combination with local, state, and private research and development investment. The federal dollar should not duplicate these investments but be used to create collaborative relationships with the non- federal R&D system so that the national effort is more productive.

Finally, we can never forget that education happens at the local level. Products like "ED-TALK," released by the Council last week just as the NAEP math achievement scores came out, synthesize the best knowledge we have so that the people on the front lines know what to do.

Mr. Chairman, we will do everything we can to aid you and the Committee in forging a federal research and development system that truly reforms America's schools.

Senator PELL. We'll now hear from Dr. Arthur Wise, president of the National Council for Accreditation of Teacher Education.

Mr. WISE. Mr. Chairman, members of the subcommittee, I am pleased to testify at this hearing on the reauthorization of OERI.

Today's schools look very much like the schools of yesterday. Meanwhile, nearly every other facet of public and private life has been radically transformed. Have you ever wondered why?

The early 20th century factory on which our schools are incidentally modelled is gone. Developments in transportation and communication have changed the way we do business. Breakthroughs in medical knowledge have transformed our lifestyles. The news and entertainment industries have been profoundly reshaped by available technology. Meanwhile, students still sit in groups of 25 to 35, presided over by one teacher, expected to progress uniformly, one grade, 1 year.

Pundits may speculate on the intransigence of the education establishment. Political leaders may despair that the schools are not ready for the 21st century. But one explanation for this failure has been left unexamined.

Investments in research and development have fueled the changes that we see everywhere in our lives. Industry in vests 4 to 7 percent of revenues in R and D. High-tech industries invest up to 20 percent and more. These investments have led to the new products and new discoveries that have changed our lives. Meanwhile, education invests a mere $100 million on a $300 billion a year enterprise, or a rate of .0003 percent, three ten-thousandths of one percent.

This is like trying to move an oceanliner with a toy tugboat.

No industry could long survive, much less improve, at the level of investment we now make in educational R and D. President Bush and the governors have set six ambitious goals for the Nation's schools. Reaching these goals will require overcoming some of the most intractable problems in American education.

As the Nation embarks on a restructuring of the educational system, it will discover that sound new knowledge and well-tested products are in short supply. It will also find that many of the proposed "solutions" to current problems have little theoretical or empirical grounding.

In short, the Nation runs the risk of perpetuating educational fadism, an affliction long plaguing our schools, where one fad gives way to another, and no real improvement takes place.

The old-fashioned factory-model school was good enough for the industrial age, which provided jobs for both skilled and unskilled workers. The information age demands an information age school which brings as many students as possible to higher order thinking skills.

President Bush is right in saying that we cannot transform education using the same strategies. Now is the time to restructure our plethora of small Federal research initiatives and place enough money into specific problem areas to effect some real breakthroughs.

It is clear that the present course of Federal research will not provide sufficient amounts of the dependable knowledge required for educational reform. What is needed is a new set of research in-

stitutes created by the Congress, a "national institutes of education". Each institute addressing an important educational goal should be funded at $50 million. Clearly, the model for the reorganization of educational research effort is the highly successful National Institutes of Health. Organized in analogous fashion, the Federal structure for educational research could help us to create a knowledge base sufficient to advance us to the goals.

One example of such an institute is the one which Congressman Owens has introduced in the House. This bill, H.R. 2467, would establish the National Institute for the Education of At-Risk Students. The institute will conduct basic and applied research on interventions likely to substantially increase the educational success of at-risk students.

The Federal Government's approach to special education already serves as a positive illustration of the potential for progress through research. Currently, the Department of Education supports the National Institute on Disability and Rehabilitation Research, funded at about $54 million per year. The import of this institute is both in its mandate and in its funding level. Enough money has been allocated to effect a profound influence on the way we educate students with disabilities.

This legislative subcommittee has an important choice to make. It can continue OERI much as it is today, in my view, assuring its irrelevance to the critical educational problems before the Nation, or it can craft legislation providing for a mission-driven structure such as has been outlined in the proposal for creation of the National Institutes of Education.

Thank you for your time and for your consideration of these ideas.

Senator PELL. Thank you very much, Dr. Wise.

[The prepared statement of Mr. Wise follows:]

PREPARED STATEMENT OF MR. WISE

(The testimony of Arthur E. Wise represents the views of the America Educational Research Association, the largest organization of educational researchers in the U.S. Dr. Wise is a former Chair of the Government Liaison Committee of the AERA. He is currently President of the National Council for Accreditation of Teacher Education.)

Mr. Chairman and Members of the Subcommittee: I am pleased to testify at this hearing on the reauthorization of the Office of Educational Research and Improvement.

Today we consider an opportunity to fundamentally alter the nation's future. Today we address one major means to reach the national education goals which President Bush has called for and which we all agree are vital. Reaching the goals will require time, effort, money and knowledge. Our base of knowledge in educational research is the foundation on which many efforts in this field should rest. But we are not yet working with a sufficient base of research knowledge that specifically addresses the problems we face today. It is time to change the structure of the federal education research and dissemination program.

My interest in this topic is long-standing. From 1973–75 I was an Associate Director of the former National Institute of Education. In the late 1970's I helped to design the U. S. Department of Education, including the OERI. Recently, I chaired the Government Liaison Committee of the American Educational Research Association which for the past three years has studied the question of how to improve the education research capacity of the nation.

The stream of events characterized by efforts to establish and achieve national goals for education, the ongoing state reform movement in education, and the powerful change potential of ideas suggested by restructuring have produced a climate

of almost desperate attentiveness to schooling and learning phenomena. There is heightened concern among a broad range of society: the business community, advocates for the disadvantaged, teachers and administrators, and the political leadership of the nation.

We agree with President Bush that sweeping, fundamental changes in our education system must be made. But the task cannot be left to school personnel or demonstration projects alone to solve. The National Goals statement calls for transformation in education. We agree that it is time for a "new educational order" in which success for all students is the preeminent national goal. President Bush's plan calls for research and development centers to create 535 "New American Schools," on which to base a renaissance in education. All of these schools will be expected to produce extraordinary gains in student learning. But without new knowledge of the education process, we are destined to repeat the mistakes of the past. With some notable exceptions we have not approached significant breakthroughs in our understanding of the problems facing our students and our schools today. Incremental change in the educational research structure, such as the creation of 535 new schools, will not do if America's leaders are truly seeking transformational change in education.

Developing "New American Schools" requires more than short-term applied research and development projects which are called for in the current plan. Creating these schools should be a part of a comprehensive approach in a federally-initiated education research program grounded on sound basic and applied educational research. This comprehensive approach has long been missing in plans for education funding.

The Need for Transformation in Educational Research

We must create new connections between what is discovered through research, the policies developed by political and school leadership, and the structures in which administrators and teachers practice. Today's knowledge base is not sufficient to answer the research questions we have, some of which have come about in the past ten years through the dramatic change in composition of our nation's students.

For example, relying on what we know today will not begin to address the tremendous need of children born addicted to crack/cocaine. According to the National Association for Perinatal Addiction Research and Education, about one out of every 10 newborns in the U.S. (about 375,000 per year) is exposed in the womb to one or more illicit drugs. The most frequent ingredient in the mix is cocaine. In major cities such as New York, Los Angeles, Detroit and Washington, many hospitals report that the percentage of newborns showing the effects of drugs is 20 percent or higher. Special schooling requirements for these children will cost significantly more than the typical per pupil expenditure. In Boston, a year of special education for a drug-exposed child can cost $13,000 compared with $5,000 per child at a regular school.

In addition to the escalating and now in many cases congenital drug problem, we are faced with other issues which challenge the limits of our knowledge. As Congressman Owens' bill H.R. 2467 cites, in many major cities the dropout rate for students is now over 50 percent. By the year 2000, an estimated 3.4 million limited English proficient school age children will be entering the school system. Teachers from the middle class will be teaching students from backgrounds vastly different from their own. Minorities will be a majority in many schools. What we know today about teaching these students is still vastly inadequate.

Much more research needs to be conducted on teaching and learning strategies. Our schools, other than a few demonstration projects and some progressive systems, are still operating under the concept of the "factory school," which taught the skills and habits needed by a newly industrialized society. Now our leaders seek to alter the mission and structure of today's schools to mirror the transformation of today's economy from an industrial to an information age. How can we work with our students to achieve higher levels of abstract reasoning ability? How should we structure our schools to deliver it?

The knowledge base we have developed thus far has made a difference in the ways we structure learning experiences for certain populations. Research has had a profound influence on how we educate students with disabilities and on young children. Prior to 1965, we focused a major share of our efforts on adolescents and middle school children. Research revealed the importance of early childhood intervention. Research on mainstreaming students with disabilities has led to better diagnosis of learning problems and improved practice. Research on pre-referral intervention strategies has led to a 25 to 50 percent reduction in the need for special education. The latter breakthrough came as a result of sustained, relatively well-funded efforts. We have not accomplished the same in other targeted areas. Why?

We have not committed ourselves to conducting the amount or type of research needed to provide some answers to our problems. Educational research has been consistently underfunded. We cannot transform schools without additional knowledge about teaching and learning. America is now at a critical juncture. Will we give ourselves a chance to transform our nation's schools, or will we simply create additional short-term demonstration projects which, history shows, have never fundamentally altered the schools?

The Current Funding Situation

Federal, state and local spending on our education system totals about $300 billion per year. Approximately $100 million is designated for education research; this is approximately .0003 percent (three ten thousandths of one percent) of the total amount of $300 billion. Some companies in private industry spend as little as five percent of their operating budgets for research and development (many companies do better). However, in one of our top industries, computer services, the top five companies spend 16 to 28 percent of their budgets on research and development (Perleman, 1989). These companies know that they must create new products and services, and that they must do it through implementing adequate research programs to stay ahead of the competition and survive in an increasingly crowded international market.

We do a much poorer job of investing in our nation's most valuable resource—our children. We have a long way to go toward even designating five percent of the total funding effort on educational research. The amount allocated for research on teaching and learning is infinitesimal; it is shameful. No industry could long survive, much less improve at the level of investment we now make in educational research and development. We have spent more on building one stealth bomber than the Secretary proposes for the entire *America 2000 Education Strategy*. The National Institutes of Health is spending one billion, five hundred fifty-four million dollars this fiscal year on cancer research. With this level of funding, we are making progress toward diagnosis and treatment of cancer. Unless we make a comparable commitment to our nation's children, we are shortchanging our nation's future and our chance to compete with other postindustrial countries on an equal basis in the twenty- first century.

Funding of education research is a historical function of the federal government and one for which it bears singular responsibility. Corporations and foundations do support some education projects and demonstration programs but they seldom support research. A National Academy of Education survey of 28 major foundations discovered that less than four percent of grants from these foundations are targeted for education research.

The federal government has, regrettably, not exercised good stewardship in this area. Indeed, over the past two decades, the federal government has been systematically disinvesting in educational research. According to the General Accounting Office, the federal investment in educational research and development declined by 33 percent in real terms between 1980 and 1987; during this same period, federal investment in research and development in all areas increased by 24 percent. (GAO, 1988). These reductions have not been accomplished without damage to the federal research program in education. In a 1987 study, the GAO emphasized the consequences of such neglect:

> "The shift away from new data collection by the research units may have serious long-term consequences for education . . . prior research may quickly lose its relevance or it may be too low in technical adequacy to sustained continued reapplication to new questions. New data must constantly be produced to meet both departmental and Congressional requirements. . . . If it is not, information will be foregone and policies will be based on less than the most complete, relevant and timely data (p. 27)."

As the nation embarks on a restructuring of the education system, it will discover that sound new knowledge and well tested products are in short supply. It will also find that many of the proposed "solutions" to current problems have little theoretical or empirical grounding. In short, it runs the risk of perpetuating educational "fadism"—an affliction long present in our schools where one fad gives way to another and no real improvement takes place.

If the marginal change approach to reauthorization is the road taken, the most that can be expected is a slightly larger amount of money for field-initiated research through the research and development centers to develop the "New American Schools." The President's plan calls for three to seven R&D teams to develop these schools. The plan states that "once the R&D is complete and the schools are launched, the operating costs of the New American Schools will be about the same

as those of conventional schools." The fallacy here is that the limited amount of applied research designated to launch the schools would be adequate for the task. For example, research on cooperative learning has shown it to be a promising approach to teaching and learning. Much more research needs to be conducted to determine how to implement this and other promising ideas on a widescale basis.

The President's plan states that "some schools may radically alter the customary modes of teaching and learning and redesign the human relationships and organizational structures of the school." The goal is one we can all support, but it will not be achieved without a long-term commitment and adequate funding resources to conduct longitudinal basic research studies. Short-term field projects are not an adequate base on which to change our entire system. The New American Schools will not solve the long-standing, underlying and intractable problems of America's schools. We must know more if we are to increase the percent of children who can function at a high cognitive level.

With the relatively large number of current research and development centers which are poorly funded for their work, a large proportion of the education research dollar goes to overhead—to travel, conference attendance, information dissemination, so that the amount left for actual research is even smaller than it appears. The 1992 budget request seeks additional funds for education but we are still operating with approximately 22 research centers funded at a total of $25 million, just a fraction over one million dollars per center. The President's plan calls for American businesses and other donors to contribute $150 to $200 million and will request an additional one-time $535 million for start up costs of the New American Schools. These expenditures will not work to further our knowledge of how to work more effectively with our school children. We must first expand our knowledge base and test ideas before we implement new strategies on a widescale basis.

Needed: Federal Leadership Through National Institutes of Education

The federal government has the clear responsibility to lead the research and development effort in education. The last significant effort was the creation of the National Institute of Education in 1971, which was initially funded at $125 million, but whose budget was quickly reduced in succeeding years so that long-term research could not be accomplished. Much pressure was brought to bear to change NIE's mission from a focus on long-term basic research to short-term projects relating to practice. In recent years the NIH and NSF also found themselves under executive branch and Congressional pressure to emphasize short-term impact at the expense of long-term inquiry. There is pressure to do something now—the "before the next election" syndrome. This is understandable, but some problems cannot be solved in one or two years. "Scientific inquiry into the educational process" was the original Congressional mandate for the NIE. In the history of education to date, scientific inquiry has had only a limited impact upon school practices. An explanation of this failure is our serious underinvestment in research into the educational process as well as a lack of sustained focus.

It is tempting to think that we can solve our problems with demonstration project after demonstration project. Many studies have documented that this approach is not sufficient. Projects can be useful, but without significant research, we are stirring the waters only enough to muddy them—and not enough to see clear results. Longitudinal research on children's growth and development may not be as politically appealing as the idea of a test to measure results. However, the results of standardized tests won't show much progress if we do not find new ways to help children learn.

You are now considering reauthorization of the Office of Educational Research and Improvement. It is a time of extreme challenge for the nation. With the changes in immigration patterns over the last 20 years, we now have a very different clientele in our schools than we did in the 1960's and 70's. The multicultural society has arrived at the same time as the information age. Higher level skills are now required to develop and service new automated technologies. At the same time, we have a dramatic increase in the number of children entering our schools whose native language is not English. The last great wave of immigration occurred as America entered the industrial age. Generations of immigrants were educated. Some reached the heights of intellectual functioning; many did not, but there were plenty of jobs for which relatively low levels of intellectual functioning sufficed. Now for the new immigrants to function, for America to flourish, and for America not to become two societies, we must find new methods of teaching and new approaches to learning. President Bush is right in saying that we cannot transform education using the same strategies. Now is the time to restructure our plethora of small federal research initiatives and place enough money into specific problem areas to effect some real findings and breakthroughs.

42-921 0 - 91 - 8

The current research structure within OERI is not mission-oriented. The Office of Research in OERI, for example, is organized by general area: education and society; schools and school professionals. This type of organization does not create a compelling set of targets for research. In addition, the current system has very limited resources. As a consequence it cannot be held accountable for results.

It is clear that the present course of federal research will not provide sufficient amounts of the dependable knowledge required for education reform. What is needed is a new set of research institutes created by Congress, which could be called the National Institutes of Education. Each institute could address an important national education problem. The institutes should be funded with budgets of $50 million each. As the title suggests, one model for such a reorganization of educational research efforts might be the National Institutes of Health (NIH). In recent years Congress has established new institutes such as the National Cancer Institute, to address specific problems.

Organized in an analogous fashion, the federal structure for education research could help us create a knowledge base sufficient to resolve problems that are presently regarded as intractable. There are many sources for deciding on the five or six mission-oriented institutes of education. Each of the six national education goals could be the focus of an institute. Virtually every objective established by the President can be seen as a call for a major research program. Another source for target areas for the institutes would be the list developed by the National Academy of Education in its latest research report.

It is important to give adequate consideration to the selection of problems that would be the central focus of each institute. However, at this stage in the reauthorization process, the most important task is to change the thinking about education research from a collection of activities designed years ago to problem-driven, mission oriented, adequately funded research institutes.

One example of such an institute is the one which Congressman Owens has introduced in the House. This bill, H. R. 2467, would establish The National Institute for the Education of At-Risk Students. The bill addresses specific problem areas: minority language issues, and rural and inner city education needs. The institute's mission is "to improve and expand the knowledge base, programs, techniques and materials concerning the educationally at-risk." The institute would conduct basic and applied research on interventions likely to substantially increase the educational success of at-risk students. It would be charged with the research and development of technology that has special merit when used with these populations, along with other specialized objectives. Monies for dissemination and evaluation would be included. The institute is mission-oriented and has specific objectives.

The federal government's approach to special education already serves as a positive illustration of the potential for progress through research. Currently, the Department of Education supports the National Institute on Disability and Rehabilitation Research, funded at about $54 million per year. The importance of this national institute is both in its mandate and its funding level. Enough money has been allocated to effect a significant amount of research. The mandate of the institute provides for a comprehensive and coordinated approach to the conduct of research, demonstration projects and related activities. The approach integrates research, development and dissemination. There is a clear focus: individuals with disabilities. The research generated through this institute is having a profound influence on the way we educate students with disabilities.

In addition to the problem-solving orientation, other features would include a headquarters for the institutes, which could coordinate activities. It might, for example, manage center and lab competitions. Headquarters would also eliminate unnecessary duplication. The mission approach would solve some of the administrative quagmires in which the current OERI has found itself. Current labs would continue their activities and support the various institutes in appropriate ways. Centers, operating at the current million or two million dollars a year, could be dedicated to one institute or could serve several institutes. With a mission on which to focus their efforts and with sufficient resources to deal with major education problems, current centers can address the objectives of each institute. In addition, each institute should set aside 15 percent of its budget to be separately managed in a field-initiated studies program so that the creativity of the field is fully tapped.

Once again, this legislative subcommittee has an important choice to make. It can continue OERI much as it is today, in my view assuring its irrelevance to the critical education problems before the nation, or it can craft legislation providing for a mission- driven structure such as has been outlined in the proposal for creation of the National Institutes of Education.

Thank you for your time and for your consideration of these ideas.

Senator PELL. Now we come to Mr. Semple, vice president and secretary of the Committee for Economic Development.

Mr. SEMPLE. My name is Nat Semple, and I am vice president and secretary for the Committee for Economic Development, and I very much appreciate the opportunity for us to testify today.

I would like to say a special hello to Senator Bingaman. Our sons were fast friends in first grade over 6 years ago, and it might be fun to bring them together and see if they recognize each other.

Senator BINGAMAN. I agree.

Mr. SEMPLE. Next year, CED will be celebrating its 50th anniversary since its inception in 1942, when President Roosevelt convened a group of business leaders to assist him in transitioning the economy from war to peace.

One of CED's earliest recommendations was the creation of the GI bill, perhaps one of the Nation's most effective educational programs. Since then, CED has made numerous recommendations that have made their way into public law including the establishment of the World Bank, the International Monetary Fund, and the Bretton Woods Agreement.

In 1968, a special CED committee which included Katherine Graham of the *Washington Post*, Pete Peterson, then president of Bell & Howell, Daniel Parker of Parker Pen, and others, released a statement which led directly to the institutionalization of the Federal role in educational R and D.

Entitled, "Innovation in Education: New Directions for the American School", the report recommended the creation of a Commission on Research, Innovation and Evaluation in Education to be established by the Congress as an independent, nongovernmental agency empowered to receive both public and private funds.

The CED was clear in what it wanted such and R and D group to do. In its research function, the CED recommended that such a commission should stimulate and encourage both basic and applied research in all branches of education; it should fund studies and authorizes improvements of curricula, and it should assess the goals of education.

In the area of innovation, CED recommended that the products of research be disseminated in a usable form to all the Nation's schools and that innovations deemed to be effective should be disseminated as widely as possible.

Finally, the CED urged that evaluation should focus on not only the basic learning skills and the acquisition of knowledge, but on the effectiveness of the schools in achieving the entire spectrum of education purposes and goals, particularly with respect to the efficiency of school administration and the effectiveness of the school as an instructional organization.

This was almost 33 years ago. And I don't mean to be a revisionist historian, but it is my understanding that this report helped spur the efforts in the Nixon Administration that led eventually to the creation of the National Institute of Education and subsequently OERI.

Now, it does not come as any surprise to me that it was a group of business executives who gave a push to a national role for educational R and D. Business leaders see R and D as a primary corporate responsibility and a key underpinning of their ability to com-

pete and prosper. This view has continued unwavering in virtually all of the reports that CED has issued on education in the last 30 years.

CED has recently clarified and expanded its view of R and D. In our 1985 statement on education reform, "Investing in our Children: Business and the Public Schools", we argued for establishing the following: a system to provide for comparative data on educational achievement. We believed at that time that we needed statistical information that would permit State by State, city by city, and school by school comparisons, and we are happy to see that this is part of the President's agenda.

We also called for regular assessments of employment readiness at least every 4 years; a continuous process of updating international comparisons, and as Senator Jeffords just mentioned, the outcomes associated with the application of new technologies.

In our most recent statement on education, "The Unfinished Agenda: A New Vision for Child Development and Education", which was chaired by Jim Renier, the CEO of Honeywell, we continued to spell out our support unequivocally for expansion of R and D. But in this report, we argue that this role of analysis needs to be expanded to include a regular analysis of those programs that help prepare children, especially disadvantaged children, for school, as well as those programs that support children once in school

In our 1987 statement, "Children in Need", we argued, and I quote, "that it was more important than ever for the Federal Government to fund high-quality research, development, evaluation and technical assistance for Chapter I, Head Start, and related programs that impact educational outcomes."

We also believe that educational researchers need to develop a new generation of compensatory education models, and school districts sorely need hands-on technical assistance from those who know how to implement and evaluate currently effective models and those that are emerging.

Although we call for new experimentation, including various approaches to school choice, we share the view of my friend on the right, that we know already what works, and what is required is not new programs but further and better dissemination of the results of those we already know function well.

I have included in my testimony the description of a school we found particularly remarkable, which is the New Futures school in Albuquerque, NM. We found this school, which dealt with pregnant teenage mothers and with disadvantaged children, to be remarkably effective—enough so that Jim Renier, the CEO of Honeywell, has established a similar program in his own corporate headquarters.

This bring me to a discussion of the President's "Education 2000" proposal. While the CED has not taken an official public position on the specifics of the President's program, we are strongly supportive of many of the objectives and concepts in the President's plan, and I have included a side-by-side analysis of CED's position with respect to the President's program, and you will see there is some substantial agreement.

I do sense, however, that there is some concern, particularly in the business community, about how the administration plans to go about implementing its proposal. One involves creation of a wholly independent new American Schools Development Corporation, with its goal of generating nearly $150-$200 million in private contributions. While many of us believe there is an important need for added R and D, there are those who believe that this should be the responsibility of the Federal Government. But then, it is certainly appreciated that given the current budget constraints in the Congress, this may be the only way to generate added support quickly.

But there are two notes of caution. First, the money that this corporation will be generating could in many cases come at the expense of other efforts already being undertaken by business. This past Tuesday we released a new study entitled, "Business Impact on Education and Child Development", where it is noted that there now exists over 140,000 separate local business-school partnerships in some 30,000 public elementary and secondary schools. Virtually all of the Fortune 500 companies are engaged directly in education efforts, efforts that reflect a considerably increased understanding of the educational process and the role business can play in improving that process.

A lot of people have a lot at stake in these efforts and will resent having their funds diverted elsewhere, and may in fact lead to a long-term disenchantment of business for further support of R and D in the future. So it is vitally important that those who take charge of the New Schools Corporation carry out their mission with these concerns firmly in mind.

A second concern is the relationship between this New Schools Corporation and existing Federal R and D efforts. It would seem senseless that OERI and the new corporation not work closely together. CED recognized the importance of join efforts over 30 years ago when it recommended an independent Federally-chartered commission where both public and private moneys could be brought together.

Finally, we hope that this R and D includes research in workable, compensatory models such as the New Futures program mentioned above.

Before closing, I just want to briefly mention—and I'm sorry I'm going beyond the time—our support for some of the other areas in the President's program including national testing and establishing a national curriculum. In the area of assessment, CED supports the establishment of national standards in the core subjects and agrees with the President that we should not develop a rigid national curriculum.

We also agree with the President's recommendation, as I have noted above, to have the National Assessment of Educational Progress conduct State by State assessments and comparisons of schools and school districts.

Finally, we agree with the President's recommendation to establish a system of voluntary examinations at various grade intervals, but we would allow the States discretion as to which grades they would involve.

We have also recommended that employers be urged to use these tests.

Mr. Chairman, business fully understands the need for R and D and the important role it plays in the future of any successful enterprise. We believe that at no time in our history have we needed a better understanding of education in this country.

Senator PELL. Excuse me for interrupting. There is a roll call vote going on, so I must ask to be excused. Please wind up, and for as long as possible one of my colleagues will preside, and if I'm not back by then, we'll recess for a moment.

Senator SIMON. [Presiding.] We thank all three of you for your testimony.

[The three publications entitled "Strategies for a New World, The Unfinished Agenda: A New Vision for Child Development and Education, and Business impact on Education and Child Development Reform," submitted by Mr. Semple are retained in the files of the committee.]

[The prepared statement of Mr. Semple (with an attachment) follows:]

PREPARED STATEMENT OF MR. SEMPLE

(The positions taken in this testimony are partially based on CED's policy statement, Innovation in Education (1968), Investing in our Children (1985), Children in Need (1987), The Unfinished Agenda (1991), and Business Impact on Education (1991). However, the views expressed herein are solely those of the author and in no way necessarily represent individual CED trustees or their organizations.)

Mr. Chairman, my name is Nat Semple and I am Vice President and Secretary of the Committee for Economic Development, an organization comprised of 250 of the nation's top business and academic leaders. Next year, CED will be celebrating its 50th anniversary since its inception in 1942, when President Roosevelt convened a group of business leaders to assist him in transitioning the economy from war to peace. One of CED's earliest recommendations was the creation of the GI bill, perhaps one of the nation's most effective educational programs. Since then, CED has made numerous recommendations that have made their way into public law, including the establishment of the World Bank, the International Monetary Fund, and the Bretton Woods agreement.

In 1968, a special CED committee, which included Katherine Graham of *The Washington Post*; Pete Peterson, then President of Bell & Howell; Daniel Parker, of Parker Pen; and others; released a statement which led directly to the institutionalizing of the federal role in educational R&D. Entitled Innovation in Education: New Directions for the American School, the report recommended the creation of a Commission on Research, Innovation, and Evaluation in Education ¶to be established by Congress as an independent, nongovernmental agency, empowered to receive both public and private funds.

The CED was clear in what it wanted such a R&D group to do: in its research function, the CED recommended that such a Commission should stimulate and encourage both basic and applied research in all branches of education; it should fund and authorize studies for the improvement of curriculum, and it should assess the goals of education. In the area of innovation, CED recommended that the products of research be disseminated in a usable form to all the nation's schools; and that innovations deemed to be effective should be disseminated as widely as possible. Finally, the CED urged that evaluation should focus on not only the basic learning skills and the acquisition of knowledge, but on the effectiveness of the schools in achieving the entire spectrum of education purposes and goals, particularly with respect to the efficiency of school administration and the effectiveness of the school as an instructional organization.

This report helped spur the efforts of your colleague, Senator Moynihan, to develop support in the Nixon Administration for establishing a locus for educational R&D, which resulted in the creation of the National Institute of Education and subsequently, OERI.

While the CED's recommendation that this be a nongovernmental agency was not adopted, the basic premise of the R&D role was.

It should not come as any surprise that it was a group of business executives who gave the big push to a national role for Educational Research. Business leaders see R&D as a primary corporate responsibility and a key underpinning of their ability to compete and prosper. This view has continued unwavering in virtually all of the reports that CED has issued on education in the last thirty years.

CED has recently clarified and expanded on its view of R&D. In our 1985 statement on education reform, *Investing in Our Children: Business and the Public Schools* we argued for establishing the following:

—a system to provide for comparative data on educational achievement. We believe we need statistical information that will permit state-by-state, city-by-city, and school-by-school comparisons; (a position that the President has recently adopted);
—regular assessments of employment readiness, at least once every four years;
—a continuous process of updating international comparisons;
—the outcomes associated with the application of new technologies to the educational process.

It was in the area of evaluating returns on investment in education where we reserved our greatest support. We called on the National Assessment of Education Progress (NAEP) to establish a regular analysis of benefits received from improving the content and quality of education.

In our most recent statement on education, *The Unfinished Agenda, A New Vision for Child Development and Education* released this past March, which was overseen by Jim Renier, the Chairman and CEO of Honeywell, we spell out our support unequivocally for expansion of R&D, and I quote:

> "Data collection and research at the federal level are crucial for continued innovation in education and children's programs, particularly in such areas as measurement and testing, pedagogy, technology use and access, and programs targeted to the disadvantaged."

Today we would argue that this role of analysis needs to be expanded to include a regular analysis of those programs that help prepare children, especially disadvantaged children for school, as well as those programs that support children once in school. In our 1987 statement, Children in Need: we argued, and I quote:

> "That it was more important than ever for the federal government to fund high-quality research, development, evaluation and technical assistance for Chapter I, Head Start and related programs that impact educational outcomes."

We believe that educational researchers need to develop a new generation of compensatory education models, and school district sorely need hands-on technical assistance from those who know how to implement and evaluate currently effective models and those that are emerging.

Although we call for new experimentation, including various approaches to school choice, we believe we know more than enough already of what works. What is required is not new programs but dissemination of the results of those we already know function well.

Let me cite just one example of a program which, if duplicated could make a significant difference in the future of children. I refer to the New Futures School in Albuquerque, New Mexico. New Futures School is an alternative school in the Albuquerque public school system, and is devoted exclusively to helping school age parents make responsible, informed decisions, complete their education, have healthy babies, and become well adjusted and self-sufficient. Since 1970 the school has offered services to over 5000 students.

This school works on the premise that teenage mothers need a good deal more support than the average student. First and foremost it creates an environment where a teenage mother is socially accepted and her individual problems are dealt with. Often an abused child herself, the teenage mom knows that she can go to a school where others share her problem. If her child is sick, she knows that he or she will be attended to by a licensed pediatrician. If she is having trouble with support payments, she will find someone who will help—all provided in a way that will enable her to attend class and to achieve her high school certificate.

The results of New Futures, longitudinally derived over 20 years, have been remarkable. Nearly 80 percent of the mothers who attend New Futures obtain their high school degree; over half go on to some form of post-secondary education. Only 16 percent end upon AFDC. And even more interesting, the repeat pregnancy rates drop by a half. This is a program known to work, and one that is being replicated—by the private sector. Honeywell, for example has established a school located at its

headquarters as part of its Success-by-Six program. In our view, the federal government needs to maintain an adequate record of these programs, and to disseminate these kinds of "success" stories.

This brings me to a discussion of the President's Education 2000 proposal and specifically with the notion of establishing an independent privately funded R&D effort.

While the CED has not taken an official public position on the specifics of the President's program, we are strongly supportive of many of the objectives and concepts in the President's plan. I would like to include for the record a more definitive comparison of CED's position with respect to the President's program, and you will see, that outside of some concern about use of vouchers for private schools, there is a good deal of agreement on the basic principles of his approach.

And again, while the CED has not taken a position on the implementation of the Presidents proposal, I do sense among many in the business community concerns about how the administration plans to go about it. One involves the creation of a wholly independent privatized New American Schools Development Corporation, with its goal of generating nearly $150-$200 million in private contributions. While I believe most see the important need for added R&D, there are those who believe this should be the responsibility of the federal government. But then, it is certainly appreciated that given the current budget constraints in the Congress, this may be the only way to generate added support quickly.

But there are two notes of caution. First, the money that this corporation will be generating could in many cases come at the expense of other efforts already being undertaken by business. This past Tuesday, we released a new study, entitled Business Impact on Education and Child Development, where it is noted that there now exist over 140,000 separate local business-school partnerships in some 30,000 public elementary and secondary schools. Virtually all of the Fortune 500 companies are engaged directly in education efforts, efforts that reflect a considerably increased understanding of the educational process and the role business can play in improving that process. A lot of people have a lot of stake in these efforts, and will resent having their funds diverted elsewhere, a resentment that might lead to a long-term disenchantment of business for further support of R&D in the future. So it is vitally important that those who take charge of the New Schools Corporation carry out their mission with these concerns in mind.

A second concern is the relationship between this New Schools Corporation and existing federal R&D efforts. It would seem senseless that OERI and the New Corporation not work closely together. CED recognized the importance of joint efforts over 30 years ago when it recommended an independent federally chartered commission where both public and private monies could be brought together. Hopefully, this relationship will be considered in the creation of this new corporation.

Finally, we hope that the R&D also includes research in workable compensatory models such as the New Futures program described above.

Before closing, I should briefly mention assessment, national testing, and establishing a national curriculum. In the area of assessment, CED supports the establishment of national standards in the core subjects, and agrees with the President that we should not develop a rigid national curriculum. We also agree with the President's recommendation, which we first made in 1985, to have the National Assessment of Educational Progress conduct state-by-state assessments and comparisons of schools and school districts.

And finally, we agree with the President's recommendation to establish a system of voluntary examinations at various grade intervals, but we would allow the States discretion as to which grades they would involve in such testing. We also have recommended that employers be urged to use these tests—which we believe should go beyond multiple choice—in their hiring decisions.

Mr. Chairman, business fully understands the need for R&D and the important role it plays in the future of any successful enterprise. We believe that at no time in our history have we needed a better understanding of education in this country. The work of OERI plays a crucial role in improving that understanding, and when combined with the President's new efforts, can help us achieve the promise of much better outcomes for the nation's youth in the future.

COMPARISON OF BUSH AND CED EDUCATION PROGRAMS

Summary

President Bush's new America 2000 Education Strategy is remarkably consistent with the body of recommendations that CED has developed in its three education reports: Investing in Our Children: Business and the Public Schools (1985), Children in Need: Investment Strategies for the Educationally Disadvantaged (1987), and The Unfinished Agenda: A New Vision for Child Development and Education (1991).

Although we may differ on some of the details, the large themes of the Bush education plan echo key CED recommendations on educational innovation, bottom-up management, increasing accountability, improving assessment, more flexible use of resources, bolstering the skills of teachers and principals, increasing parental involvement, identifying employability skills, and recognizing the importance of the "social agenda" and the need to go beyond narrow school-based reform. Nevertheless, the Bush education program contains few specific early intervention and early childhood education strategies to address the number one national education goal – Readiness to Learn. However, some initiatives currently underway at the Department of Health and Human Services are addressing this issue. The following compares specific points in the America 2000 Education Strategy with CED's policy positions.

Bush Education Program	CED Policies on Education and Child Development
Standards and Assessment	
National standards in five core subjects (English, math, science, history, and geography) will be established. The President does not support a national curriculum.	In Investing in Our Children, CED called for all children to master the basics of written and verbal communication, math computation and reasoning, and have a common knowledge of history and literature. CED also called for the development of a "common curriculum" that incorporates the skills and knowledge needed in a modern, technological society. CED has not supported the development of a rigid national curriculum.
The President wants to allow the National Assessment of Educational Progress (NAEP) to conduct state-by-state assessments and comparisons of schools and school districts. Schools, school districts, and states will be encouraged to issue regular report cards on their educational performance.	CED made this recommendation in Investing in Our Children and again in The Unfinished Agenda. CED has also called for comparative reporting by schools, school districts, and states.

A system of voluntary national examinations, which will test both basics and higher order skills, will be developed for fourth, eighth and twelfth grade students in the core subjects. Employers will be urged to pay attention to the tests in hiring.

Investing in Our Children recommended achievement testing at specified intervals (i.e. third, fifth, or eighth grades) in basic reading, writing, and math skills as well as in key content areas. States would have the discretion on which grades to test. CED has also consistently called for the development of new assessment mechanisms that go beyond multiple choice tests to assess higher order skills, i.e. performance assessments or portfolios. In The Unfinished Agenda, we recommended that employers asked prospective entry-level employees for high school transcripts or report cards.

Innovation

The Bush program calls on business to lead and fund a new nonprofit research and development corporation that will identify and create innovative approaches to education. Business will be asked to raise $150 to $200 million for this effort.

Part of the Bush innovation effort involves the creation on 535 new "America 2000 Schools." Each school will receive $1 million in start-up funds to apply innovative approaches and become national models of innovation.

In all three of its policy statements on education, CED calls for more federal support for research and development in innovative approaches to teaching and learning for all students, particularly the disadvantaged. CED has also consistently called on local and state education systems to apply the principles of programs that work in a systematic way.

Choice

The Bush plan supports increased parental choice as a primary strategy for improving schools. The centerpiece of the Bush plan is a $200 million incentive grant program to encourage local communities to adopt choice plans. It is understood that to qualify for a grant a community would have to include private schools in their choice plan. Bush also proposes a $30 million national school choice demonstration project and allowing Chapter 1 money to follow individual children to whatever school they attend.

In The Unfinished Agenda, CED supports choice only in the public schools and where it is part of an overall program of educational restructuring that pays particular attention to the needs of the disadvantaged. CED also supported public school choice in Investing in Our Children through a system of "universal magnet schools" that would ensure a broad range of effective educational programs for all children in a community.

BEST COPY AVAILABLE

Teachers, Principals, and School Site Reform

The President's plan focuses on individual schools as the locus of change and proposes more local autonomy for teachers, principals, and parents to decide how the school should operate.	This proposal directly parallels CED's consistent call for "bottom-up" reform that propels change in the school building and the classroom.
The President proposes greater flexibility in the use of federal resources for education in exchange for enhanced accountability for results.	CED made this recommendation in The Unfinished Agenda. Earlier, in Investing in Our Children CED called on states to exempt the best performing school districts from unnecessary rules and regulations.
The President would provide federal seed money for professional academies to upgrade the leadership skills of principals and teachers in each state.	This proposal is consistent with CED's calls for the revitalization of Teachers Centers and programs to improve principals' management skills in Investing in Our Children.
The Bush plan would encourage differential pay for teachers of core subjects, who teach in difficult situations, who mentor other teachers, and who teach well.	This is very consistent with recommendations in Investing in Our Children on using pay differentials and other financial incentives to reward excellent performance and to attract qualified teachers to shortage areas, such as science, math, and inner-city schools. Also recommended were new teacher career ladders, which would include mentoring roles, that would tie compensation to greater responsibility. CED also recommended small grants and fellowships to enable teachers to develop and replicate innovations in the classroom.
The Bush plan would provide grants to encourage states and local districts to develop alternative certification routes for teachers and principals.	Investing in Our Children recommended greater flexibility in certification requirements, but goes much further in proposing improvements overall in teacher training and development.

BEST COPY AVAILABLE

Incentives

The President has proposed $40 million in grants to school districts that show significant gains in student achievement and scholarships to reward academic excellence among low income students. Citations will be given to high school students who excel on the new achievement tests.

CED has made no recommendation on the issue of monetary rewards to individual students, but in Investing in Our Children we endorsed the concept of awarding honors diplomas to reward students for superior achievement. CED also recommended giving financial awards to an entire school staff to reward overall school improvement.

The Social Agenda

The President calls on communities to become "America 2000 Communities," by agreeing to adopt the six national education goals, developing strategies for meeting the goals, and measuring their attainment. Each America 2000 community must be willing to create one of the 535 New American Schools (see above). Priority will be given to communities with large numbers of at-risk children.

In spirit, this proposal corresponds to CED call in The Unfinished Agenda for broad-based communitywide coalitions of business, education, and community leaders to tackle change at the local level. A key CED recommendation calls on communities and their school systems to conduct an inventory of how they are meeting the educational and developmental needs of children and to devise strategies for addressing unmet needs.

The Bush program calls on the nation's governors and the Domestic Policy Council's Economic Empowerment Task Force to develop strategies for streamlining and coordinating federal, state, and local health and human services programs for children and families. At the federal level, the Department of Health and Human Services recently announced the consolidation of a number of separate programs for children and families, including Head Start, under a single children's agency. HHS and the Department of Education have also established an interagency task force to develop strategies for linking schools with social services.

In The Unfinished Agenda, CED calls for greater coordination of programs and new linkages between health and human services program and the education system at every level.

Links to the Work Place

Business and labor are being asked to establish job-related and industry-specific skill standards and to develop skill certificates.

In Investing in Our Children we call on employers to provide regular feedback to schools on the skill needs of the work place. In The Unfinished Agenda CED urges business to take a leadership role in working with education and the community to develop performance-based goals that reflect real adult skill needs.

The President's proposal calls for the development of Skill Clinics in every community to help current workers upgrade their skills and knowledge.

This strategy is generally supported in An America That Works, which identifies effective examples of consortium-led industry-based training programs.

Senator SIMON. If I may ask Dr. Nowakowski and Dr. Wise, is there any possibility of your ideas meshing? Is there a conflict in what you are suggesting, Dr. Wise, with what Dr. Nowakowski is suggesting?

Mr. WISE. I wouldn't say so, Senator. I would say that the concept of the national institutes of education as I explained them is meant to embrace the existence of centers and laboratories, and certainly we agree in the sense that existing centers are not well enough financed to conduct sufficient research.

I would advocate expanding the budgets of existing ones rather than reducing the number of existing ones, but that would be a minor difference between us.

Ms. NOWAKOWSKI. I would agree, Senator. I think we have about a $200 million investment in R and D, and I think we need about a $2 billion investment. And I think Dr. Wise and I agree that we must have a critical mass, organizations that have a critical mass of 50-200 researchers focused on major social problems in order to make a big difference.

Senator SIMON. And as I follow what you are suggesting, Dr. Wise—and incidentally, let me note that your association, the National Council for Accreditation of Teacher Education, stole a very valuable member of my staff, Judy White, so your testimony gets great weight in our office, I want you to know—but you are suggesting that we focus in a little more, just as the National Institutes of Health focus on arthritis or diabetes or cancer.

Mr. WISE. I think it is very important, sir, that we have very clear targets for Federal educational research and development. Having the highly fractionated approach that we have followed for the last couple of decades has really not gotten us very far, and creating a small number of well-financed institutes that would work on some of the most intractable problems of American education over a long period of time will, I think, do more for helping us to realize the aspirations of the governors and the President than the kinds of approaches that I see being espoused by lots of people these days.

Senator SIMON. Mr. Semple, I read with great interest the example of the New Futures school in Albuquerque, NM—and I see our friend Senator Bingaman has left already—but that is an exciting example. When you say, "although we call for new experimentation, including various approaches as school choice, we believe we know more than enough already of what works. What is required is not new programs, but dissemination of the results." When you say, "we know more than enough already of what works", I guess I would differ slightly with you there. I think research has to continue. If General Motors or anyone else says, "We have the ultimate knowledge, and we are going to stop experimenting and stop doing research"——

Mr. SEMPLE. I may have misstated myself. What I really mean to say is that we know enough now to start doing things, and there are things we can do now that we are not going because we know there are things that work; but that doesn't mean we should stop R & D. No question about it. I am sorry if I misstated myself.

Senator SIMON. OK. We are in agreement.

Senator Kassebaum.

Senator KASSEBAUM. [Presiding.] Thank you, and I'll take only a minute because time is running out on this vote.

Both Dr. Wise and Dr. Nowakowski made somewhat similar comments regarding, say, a "super institute" or a "national institutes of education." Dr. Nowakowski, you mentioned Congress defining a research agenda.

I would suggest, as a matter of fact, that Congress would certainly have a role, but that should come from the Department of Education. It seems to me they are the ones who should define the agenda. We obviously always have a part in commenting on such, but I think Congress defining it is perhaps not the best place to start. That is a small point. I would gather that what the Secretary was saying is something you would agree with, that there should be more of a focus on the issues. That would help in the dissemination of information. Research having a sharper focus would then be regarded perhaps with greater esteem, so to speak, and the funds targeted in a better way.

Would you accept that?

Ms. NOWAKOWSKI. Senator, I think perhaps this kind of a collaboration may need to take place. The Secretary talked about much of the information generates. We do so in our region, about our region and who we are and what is going on, and I think that needs to inform our Nation's priorities. But also we need to make sure that the titles of R and D centers don't change every 5 years. Our R and D centers, 29 of them, are spread across 60 universities. Some of them have $900,000 budgets and no more than one or two full-time staff, and they are dealing with issues like professional accountability and leadership and reading. And if Congress can help collaboratively focus on the Nation's major issues and give that kind of direction and reinforcement to the Department of Education, we have a chance of focusing more capacity and certainly more long-term stability on solving some of these problems with a critical mass.

Senator KASSEBAUM. I guess I would rather see us improve what we have than to start into the creation of "super institutes" at this point. I am just one who, when we talk about that kind of money, would give anything to see that going to teachers who are teaching in underserved areas, teaching the children at risk. Research is very important, but when we do have limited resources, if we could give that to teachers who are working in areas where we need to implement some of the research, I think that that could be a very valuable use of funds.

Ms. NOWAKOWSKI. Well, it is sad that we have to make decisions. I wish we could have them both.

Senator KASSEBAUM. Yes, that's what we always wish in these instances.

Mr. WISE. If I may say, Senator, sometimes the role of educational research and development is underestimated in the sense that an investment in research and development can have much more powerful consequences sometimes than direct expenditures at the local school level.

At the beginning of my testimony I suggested that there is a high level of dissatisfaction in the country with the nature and quality of schooling, and schools look pretty much like they used to

decades ago, whereas every other facet of our lives have changed over the decades quite dramatically. I attribute that directly to the fact that we invest 5 to 10 to 20 percent of operating revenues on research and development in all these other sectors. And somehow when it comes to education, we think we can reform the system on a pittance, and I'm just afraid that that is not realistic.

Senator KASSEBAUM. Mr. Semple, I'm going to have to go and vote, but I did read the report and recommendation that the council did about children, and I thought it was a very interesting report. I think, as you say, dissemination of some of this information and how that is handled could be very useful.

Senator Pell will submit questions to this panel. When we return from the vote, we will convene the second panel. We thank you very much and regret that we have had this interruption.

Thank you.

[Recess.]

Senator BINGAMAN. [Presiding.] I wanted to ask Dr. Wise a question, if I could, before we start the next panel, and I apologize for fouling up the order of things, and maybe this was asked after I left the room. My question is whether the suggestion by the President that we have these 535 new schools helps or hurts the situation.

It sounds to me from your basic statement that you think we've got this research activity way to dispersed already and that it is not working to help us actually upgrade the quality of education. And as I understand it, the main component of what the President has composed with regard to research, other than increased assessment, involves these new schools.

If you'd be willing to comment on that, I would appreciate it.

Mr. WISE. Well, the history of the demonstration school approach to educational reform has been shown not to work. We had a corresponding effort in the mid-Sixties known as the "Experimental Schools Project" for some $60 million, which was spent and I suppose made life more pleasant in those schools which received grants under that program, but that $60 million disappeared without any lasting trace on American education.

Similarly, also during the Sixties, the Ford Foundation invested very heavily in the demonstration school approach to trying to improve American education, and they spent millions of dollars, at least, trying to reform education by having model or demonstration schools, each trying to improve themselves.

Subsequently, the Ford Foundation commissioned a research report which was called "A Foundation Goes to School" in which is kind of contritely admitted that the approach which it had taken had failed in its effect, which was to transform American education back then.

I see the $535 million school operation in quite the same way. I think it will create short-term excitement and good education in those schools which happen to be blessed in the competition to become one of those. But in terms of producing reliable knowledge that will transform every American school the way we want to see them transformed, I think that is not a cost-effective use of Federal money. I believe a far more cost-effective use of Federal money is systematic research and development, more essentially managed.

I said at the beginning of my statement that the scale of investment in R and D is minuscule in relation to the magnitude of the enterprise of K through 12 education.

If you really want to move this monster, you really have to start using the strategies which we have used to transform every other sector of public and private life, which is systematic research and development. That is why our homes look different today, our hospitals look different today, transportation looks different today, the way we do business, our offices, all look entirely different from how they looked 50 or 75 years ago. Meanwhile, our schools look the same, and I attribute that to the minuscule level of effort that we make trying to improve our schools.

Senator BINGAMAN. So in your view, if we were to organize ourselves around these mission-driven institutes and fund them at an adequate level and then try to have a national system for disseminating the results of that research, that would go a lot further toward actually upgrading the quality of education.

Mr. WISE. I believe so, sir. I think if you kind of look at the history of the National Institutes of Health since the mid-Forties or mid-Fifties and the way in which we have mounted systematic research and development campaigns under the auspices of each one of the individual institutes, we can see how we have made major inroads in the practice of medicine and the treatment and prevention of various diseases. I believe that we must take an analogous approach in this arena.

Senator BINGAMAN. OK. Thank you very much. I appreciate that.

We'll call the next panel, and I'll continue until the chairman returns. If Mr. Boehlje and Dr. Stewart are still here, we'd appreciate them coming up.

Mr. Boehlje, why don't you go ahead with your testimony, and I'll try to learn enough for the entire committee.

STATEMENTS OF BOYD W. BOEHLJE, BOARD MEMBER, NATIONAL ASSESSMENT GOVERNING BOARD, WASHINGTON, DC: AND DONALD M. STEWART, PRESIDENT, THE COLLEGE BOARD, WASHINGTON, DC

Mr. BOEHLJE. Thank you.

I appreciate this opportunity to testify about the National Assessment of Educational Progress, or NAEP. I should give you some background on myself. My name is Boyd Boehlje. I was appointed to the National Assessment Governing Board in August of 1990. I am also a member of the Pella, IA school board, and that is the slot that I fill on the National Assessment Governing Board. I am also an officer of the National School Boards Association.

My comments today reflect the position of the board of the National Assessment Governing Board.

The function of the board is to set the policy for NAEP. NAEP is our only national representative assessment of America's students and what they know and what they can do. I'd like to highlight this morning four board recommendations that bear on reauthorization of NAEP. Those relate to State-level reporting, use of

NAEP for below State reporting, conducting NAEP annually, and the independence of the board.

My written testimony provides additional background on these and other recommendations that may require legislation.

Relating first to State representative reporting, the board believes that the NAEP legislation should be amended to provide a general authority to conduct regular State representative assessments. Congress authorized NAEP trial State assessments in 8th grade mathematics in 1990, and in 4th grade reading and 4th and 8th grade mathematics in 1992.

At this point, 40 jurisdictions, including 37 States, Guam, Virgin Islands and the District of Columbia, have already participated in the 1990 trial; 43 States have already signed up for the 1992 State trial assessments.

On June 6th, the first ever State comparable data were released for 8th grade only from the 1990 math assessment, along with national data for grades 4, 8 and 12. The indications so far are that the data are being reported appropriately, and the States are finding the information useful in helping to find ways to improve student achievement.

A congressionally-required evaluation by the National Academy of Education recently recommended expanding the trial to three grades and three subjects in 1994. There is no State-level testing under NAEP authorized beyond 1992, yet our initial experience seems to be that the trial is working very well at this point. The board hopes that Congress will give favorable consideration to this recommendation.

In regard to——

Senator BINGAMAN. Let me ask, the recommendation being that we expand it to three levels—is that the recommendation you want us to give favorable consideration to?

Mr. BOEHLJE. The recommendation is basically that there be a general authority to conduct regular State representative assessments, and yes, most likely at three levels. But at this point the authority was granted only for a limited period of time and in a limited area, apparently as a test to see whether it is going to be effective or not.

The board further believes that NAEP legislation should be amended to permit States, school districts and nonpublic schools to use NAEP at their option and at their cost to collect data and report on performance of schools and school districts. This option was available and used prior to 1988 and continues to be desired by many States and districts. Under the current law, it is prohibited. Today a district or school that wants to cannot use NAEP to compare its results with its State or other similar districts, its region or the country, or to monitor its own progress over time.

We recognize that some organizations such as NSBA oppose the local use of NAEP. The concerns are that it will permit unwarranted intrusion into local decisionmaking and will tend to establish a national curriculum. Those concerns are also shared by the National Assessment Governing Board. But it is the position of the National Assessment Governing Board that those concerns would not result in the actual establishment of a national curriculum. There would have to be several conditions to be present for that to be ef-

fective or to happen. There must be consequence as a result from the testing, there must be an entity with jurisdiction to set out and enforce the curriculum requirements. None of those conditions are present within NAEP nor are they intended under the board's recommendation.

The board believes that a permissive authority should be granted to allow annual testing and replace the current mandated biennial cycle. This would provide NAEP with the flexibility to be responsive to information needs as they evolve. Annual NAEP testing and reporting are needed to provide timely and sufficient data for policymakers and the public, and for reporting on progress toward National Education Goal 3 for student achievement.

These items were all adopted by the board in December of 1989. They are embodied in the NAEP amendments described this morning by Secretary Alexander. They are incorporated in the President's "America 2000: Excellence in Education Act" introduced by Senators Kennedy, Pell, Hatch and Kassebaum. The amendments to NAEP would provide regular State assessments beginning in 1992, would require annual data collection and reporting for NAEP, and remove the prohibition against reporting school and school district results.

One other issue I think is important to bring to your attention. The board believes that with respect to its role and responsibilities, the NAEP legislation contains ambiguities and conflicting provisions. For example, the NAEP legislation requires the board to exercise its functions, powers and duties independent of the Secretary and the other offices and officers of the Department of Education. At the same time the legislation places the board within the department. This has raised a number of procedural and operational issues for the board and the department.

There is additional detail in my written testimony in regard to this issue.

At the present time, the board and the department are attempting to see whether, through a memorandum of understanding, they can reduce the ambiguities and contradictions, but that may or may not be possible.

Thank you for the opportunity to present this information.

Senator BINGAMAN. Thank you.

[The prepared statement of Mr. Boehlje (with attachments) follows:]

PREPARED STATEMENT OF MR. BOEHLJE

Mr. Chairman, let me express my appreciation for the opportunity to testify before the Subcommittee on Education, Arts and Humanities about the National Assessment of Educational Progress (NAEP).

My name is Boyd W. Boehlje. I was appointed to the National Assessment Governing Board (the Board), in August of 1990. The Board sets policy for NAEP, administers the national consensus process that determines the content for each NAEP test, and identifies achievement goals for each grade and subject tested under NAEP. I am a member of the Board's Ad Hoc Committee on Reauthorization and its Math and Science Committee. I am also a member of the Pella, Iowa school board and an officer of the National School Boards Association. By profession, I am an attorney in private practice.

NAEP, in its twenty-two year history has developed a well deserved reputation for integrity and innovation. Now known as "The Nation's Report Card," NAEP is our only nationally representative assessment of what America's students know and can

do. On June 6, 1991, as part of a pilot project authorized by the Congress in 1988, NAEP reported for the first time State-representative results from the 1990 mathematics assessment.

I am honored to provide testimony on this esteemed federal program. However, I would want to emphasize that my remarks about possible changes to the NAEP legislation will be limited to policies adopted by the Board. As you know, the NAEP legislation specifies the responsibilities and functions of the Board. The law also stipulates that ". . . the Board . . . shall be independent of the Secretary (of Education) and the other offices and officers of the Department of Education." Thus, my remarks will not necessarily represent positions of the Department of Education.

This morning I will provide an overview of NAEP and the role of the Board, discuss the utility of State-representative data, and describe Board policies that require legislative action.

Overview of NAEP

The purpose of NAEP is ". . . to improve our Nation's schools by making objective information about student performance in selected learning areas available to policymakers at the national, regional, state and local levels."

Under current law, the National Assessment is conducted every two years. The subjects covered and their prescribed cycles are: reading and mathematics, at least once every two years; writing and science, at least once every four years; and history/geography at least once every six years. The law also permits the Board to select other subject areas for assessment, including assessments of adult literacy.

The National Assessment reports results on students at ages nine, thirteen and seventeen and in grades four, eight and twelve. However, it is important to point out that minimizing test burden is an overriding NAEP policy. NAEP does not test all students at these ages and grades, it tests only a small sample. For example, of the more than nine million students in the three grades covered by NAEP, approximately 51,000 participated in the newly developed 1990 reading assessment and approximately 2,500 students per participating State took the 1990 eighth grade mathematics State trial assessment.

Another way NAEP limits test burden is by minimizing student testing time. Testing time under NAEP is about one hour per tested student. This is accomplished by a procedure known as matrix sampling. In matrix sampling, equivalent subsamples of students receive different portions of the total pool of test questions. The total student sample is large enough to estimate precise results for each question. Since NAEP is intended to report aggregate student results at the national and regional levels and, under the trial State assessment, at the State level, this is an efficient way to assess student performance.

One final aspect of NAEP should be highlighted—its capacity both to report trends over time and to incorporate new developments in curriculum and in assessment technology. To do this, NAEP conducts parallel assessments, one for trend data and one that is "cross-sectional," to reflect current thinking and practices. For trends, NAEP uses procedures and items from past assessments; this maintains what is now almost a twenty-year trend line in subjects such as reading and mathematics. The cross-sectional assessment reflects current research and best practice in each subject area identified through a national consensus process that is administered by the Board. I will be talking more about this consensus process later in my testimony.

ROLE OF THE NATIONAL ASSESSMENT GOVERNING BOARD

In 1988 the Congress reauthorized NAEP, making improvements, providing for the State trial assessments, requiring evaluations, and redesigning the governance structure to include an independent Board to formulate policy for NAEP.

Prior to 1988, NAEP was conducted under a grant from the Department of Education. The pre-1988 NAEP authority required the grantee to establish an Assessment Policy Committee (APC) to advise on policy matters. This arrangement was intended to insulate NAEP policy development and the conduct of the assessment from the Department of Education. But it also had the effect of impeding the Department's ability to assure accountability in the administration of the NAEP program.

Under the 1988 NAEP reauthorization, the Congress created a tripartite structure for NAEP: an independent governing board to formulate policy for the assessment; administration of the NAEP program within the National Center for Education Statistics; and conduct of NAEP through contracts, grants or cooperative agreements. The result is a system of checks and balances that provides for independence in NAEP policy development, permits the Department of Education to remain at

"arms length" from testing policy and allows appropriate accountability for program administration.

The legislative history of the current NAEP legislation indicates that the independence of the Board was an important consideration; both the Senate and Conference reports discuss provisions that have this purpose. At least three provisions of the law appear to be directed at the independence of the Board; specifically, the provision already mentioned regarding the Board's independence of the Secretary; the requirement that the Board "exercise its independent judgment free from inappropriate influences and special interests"; and the requirement that appointments to the Board are made by the Secretary of Education only from candidates nominated by the Board.

In addition, the composition of the Board as prescribed in the law seems to be founded on the idea of an independent Board. The Board has twenty-three voting members: two governors and two state legislators (both on a bipartisan basis); three teachers; an elementary and a secondary school principal; a local and a State school board member; two chief state school officers; one school superintendent; two curriculum specialists; two testing experts; one nonpublic school administrator; a representative of business or industry; and three representatives of the general public. The only federal member, the Assistant Secretary for Educational Research and Improvement, serves in a non-voting capacity.

It is difficult to imagine a group of individuals who, by the nature of their private roles, would be more grounded in the principle of State and local primacy in education matters and less inclined toward federal interventionism.

We believe that a 1987 report about NAEP was considered by the Congress when it took up the 1988 NAEP reauthorization. The report of the Study Group headed by then Governor (now Secretary of Education) Lamar Alexander and H. Thomas James—entitled "The Nation's Report Card"—recommended a tripartite governance structure similar to that in current law. With respect to the body that would be responsible for NAEP policy, the Study Group recommended that:

The governance and policy direction of the national assessment should be furnished by a broadly representative (Board) that provides wisdom, stability and continuity; that is charged with meshing the assessment needs of states and localities with that of the nation; that is accountable to the public—and to the federal government—for stewardship of this important activity; but that is itself buffered from manipulation by any individual, level of government, or special interest within the field of education.

With this as background, I would like to discuss some of the specific responsibilities of the Board as set out in law.

Formulating Policy

The general responsibility of the Board is to formulate the policy guidelines for the National Assessment. This includes formulating policy for: the methodology of the assessment; analyzing data and reporting results; selecting non-mandated subjects for assessment; and improving the form and use of the National Assessment.

In addition, the Board has responsibility for two activities that are relatively large in scope and essential to the conduct of NAEP. The first of these, developing consensus on the content for each assessment, forms the foundation for each assessment. The second, identifying achievement levels for each grade for each assessment, establishes benchmarks for interpreting the results.

Consensus Process on Test Content

Developing consensus on the content of each assessment involves much work. With a decentralized public education system of 16,000 school districts and fifty-odd state education agencies (including the District of Columbia and the territories), the need to include non-public schools, and the importance of incorporating the perspectives of researchers, practitioners, and subject area specialists, it becomes quickly apparent that arriving at a consensus in each subject area on what should be included in NAEP is an immense and complicated undertaking. The process involves literally hundreds of people over a period of twelve to eighteen months at a cost of approximately $450,000-$550,000 per subject area.

The result of this consensus activity is a document that, for each grade, describes in very specific detail what the test will include. This document is used by the NAEP contractor to develop the test questions that will be used for the assessment in that subject area.

Achievement Levels

Setting achievement levels for each assessment is also a large and complicated undertaking, and at the present time, the subject of intense scrutiny. Setting achievement levels, or defining what students should know and be able to do in each grade and subject tested under NAEP, was required for the first time under the 1988 NAEP reauthorization. Unlike consensus on test content, setting achievement levels has never previously been done.

Realizing that it was in uncharted territory, the Board has attempted to proceed prudently. In May 1990, the Board decided that its first effort would be a trial limited to one subject—in this case mathematics—as part of the 1990 assessment. The Board has obtained the advice of the foremost experts in the field of education standard-setting to design the methodology for setting achievement levels. The actual work of setting the achievement levels involves the participation of both educators and non-educators. In addition, the Board has conducted three public hearings to receive comment at various points in the process.

The Board conducted a validation/replication study of the initial work on achievement levels for the 1990 mathematics assessment. a Concurrent with these activities is an external evaluation of the process. The evaluation will continue through the reporting of the 1990 NAEP mathematics results using achievement levels to assess their utility in interpreting NAEP results. Overall, setting achievement levels will involve hundreds of individuals over a period of about eighteen months at a cost of approximately $350,000.

The result of this first achievement level-setting activity is a document that describes what students should know and be able to do in mathematics at the fourth, eighth and twelfth grades that will be used for interpreting results on the 1990 mathematics assessment. Called for in the legislation, we intend to build upon this experience so that this ambitious trial effort can be improved in subsequent assessments.

NAEP REAUTHORIZATION ISSUES

Over the last eighteen months, the Board has adopted a number of recommendations that would require legislative action. These recommendations and the rationale for each are described in detail at attachment A. I would like to highlight four of those policies here: state-representative reporting, use of NAEP for below-state reporting, conducting NAEP annually rather than bi-annually, and the independence of the Board.

The first two address a common theme: use of NAEP for reporting results at the State and below the State (i.e. district and school) level. The Board views these two issues differently, and I will discuss each in some detail below. However, as a general principle, the Board believes that State-representative and school- and district-representative results, properly used, can be very helpful in informing policy and in improving instruction.

For example, in addition to average overall scores for the 1990 eighth grade mathematics State trial assessment, there will be subscores by content area (e.g. numbers and operations, measurement, geometry, data analysis, and algebra). In addition, background questions in areas such as teacher certification, in-service training, instructional practices, and parental involvement are also part of the assessment.

Thus, a State superintendent of instruction might compare State results on the subscales with those of a similar or neighboring State, and at the same time look at teacher certification and in-service training data. This kind of comparison can help identify strengths as well as weaknesses, identify States that may have useful models, but most importantly, it would provide information on which some action can be taken that would be aimed at improving the education received by the students.

The Board is not in favor of comparisons for comparison's sake, and has adopted policies for reporting State results that would avoid the inclination to misuse or misinterpret results.

1. *Authorizing State-representative assessments.*

Congress authorized NAEP trial State assessments in eighth grade mathematics in 1990 and in fourth grade reading and fourth and eighth grade mathematics in 1992. No subsequent State testing is authorized. These trial State assessments are being closely monitored by the Department of Education, the NAEP contractor and the Board and will be the subject of analysis across the country. A congressionally required independent evaluation is being conducted by the National Academy of Education; an interim report was delivered on April 1, 1991 and additional reports are expected in September 1991 and in late 1993.

On June 6, 1991, the National Center for Education Statistics released data from the 1990 State trial assessment in eighth grade mathematics, along with national results for grades four, eight and twelve. Thirty-seven States participated in the 1990 trial and forty-three States have already signed up for the 1992 State trial assessments. Unless the evaluation proves that the State trials do not yield valid, fair and accurate data, and unless the State interest in participating evaporates, the groundswell of support for regular State-representative data as a part of NAEP is likely to continue.

The Board believes that, given the favorable results from the interim evaluation, the NAEP legislation should be amended to provide a general authority to conduct State representative assessments, with flexibility to determine the frequency and subjects to be assessed, with the full costs borne by the federal government and with State participation voluntary.

2. *Restoring local options to use NAEP*

The Board believes that the NAEP legislation should be amended to permit States, school districts and non-public schools to use NAEP, at their option and cost, to collect data and report on performance of schools and school districts.

Under current law, ". . . information [collected under NAEP] with respect to individual schools [shall] remain confidential . . ." The law further provides that "The use of National Assessment test items and test data employed in [the trial state assessment] to rank, compare, or otherwise evaluate individual students, schools, or school districts is prohibited." Eliminating the prohibitions against local use of NAEP, with respect to school and district reporting, would restore the option that existed prior to the 1988 reauthorization of NAEP. This option was exercised by many states and districts, and continues to be desired by many states and districts.

It is not the policy of the Board that reporting below the State level should be a federal activity. The Board has never suggested—nor does it now—that the organization that conducts NAEP should establish samples in individual schools and school districts as part of the NAEP contract and report the results. The Board views this as a local control issue—just as no State or district should be required to participate in NAEP, neither should a State or district be prevented from using NAEP to assess and report aggregate student performance in its schools.

It should be noted that some education organizations are opposed to the position that states, school boards and non-public schools should be permitted to use NAEP for reporting school and district results. The primary concern of these organizations is that this use of NAEP will permit unwarranted intrusion into local decision making and will tend to establish a national curriculum.

The Board is concerned about these issues as well. The Board's recommendation is compatible with these positions in the following ways:

1. Local use of NAEP would only occur at local initiative; the decision, costs, administration and reporting would all be under local control. There would be no connection between local use of NAEP and other federal funding. Since the decision would be a local one, there could be no federal intrusion.

2. In order to establish a national curriculum, several conditions must be present: students must be tested on a regular basis, there must be consequences that result from testing, and there must be an entity with jurisdiction to set and enforce curriculum requirements. None of these conditions is present within NAEP, nor are they intended under the Board's recommendation. Local use of NAEP as proposed by the Board would allow state and local education authorities to do as they always have done in determining curriculum, instruction and funding matters.

It is important to the use of NAEP and its future value as an assessment tool that it not be used as a mechanism to determine qualification for school funding or to determine curriculum to be taught at local schools. It is equally important that such a valuable resource be utilized most effectively to accurately reflect the state of educational progress being made by the students at different grade levels and in various circumstances to aid each school in their own evaluation of their education progress.

3. *Conducting NAEP on an annual rather than biennial schedule.*

The Board believes it would be better to give NAEP on a rotating schedule of three subjects per year rather than the current practice of testing three to five subjects every two years. Annual NAEP testing and reporting are necessary to provide timely and sufficient data for policy-makers and the public. This would allow NAEP to provide key measures of academic achievement for the annual progress reports on American education, called for last year at the Education Summit between President Bush and the nation's Governors.

An annual NAEP would produce little net increase in test burden because the current every-other-year cycle includes up to five tests each testing year and exten-

sive pre-testing in off-years. NAEP conducted annual testing from 1969 until 1980, so this proposal is not without precedent.

The Board believes a permissive authority allowing annual testing rather than a mandated biennial cycle would provide the flexibility for NAEP to be responsive to information needs as they evolve. Of course, the number of subjects actually assessed would be subject to the availability of appropriations.

I would be remiss if I did not point out that President Bush has submitted legislation to the Congress that would accomplish these three objectives. The "America 2000 Excellence in Education Act," introduced by Senators Kennedy, Pell, Hatch and Kassebaum in the Senate and Representative Goodling in the House, contains amendments to NAEP that would require regular state assessments (for states that choose to participate) beginning in 1992, remove the prohibition against reporting school and school district results, and require annual data collection and reporting for NAEP. These provisions are consistent with recommendations of the Board, and we would hope that the Congress will be able to act on them quickly.

4. *Clarifying the role of the Board.*

The Board believes that, with respect to its role and responsibilities, the NAEP legislation contains ambiguities and conflicting provisions. For example, the NAEP legislation gives the Board the authority to "formulate the policy guidelines for the National Assessment," to "exercise its functions, powers and duties independent of the Secretary and the other offices and officers of the Department of Education," to "exercise its independent judgment, free from inappropriate influences and special interests," and to "hire its own staff."

Taken together, these provisions could not be more clear in their intent to establish an independent policy body for NAEP insulated from the jurisdiction of the Department of Education. However, other provisions reduce this clarity.

Under the NAEP legislation, the Commissioner of Education Statistics carries out NAEP "[w]ith the advice of the National Assessment Governing Board [emphasis added]." This has been interpreted at times by the Department of Education to mean that the Board is an advisory committee whose policy decisions must only be given consideration by the Secretary. Another part of the law provides that "The Secretary may appoint, at the direction of the Board" up to six technical employees under an excepted service appointing authority. This provision has the potential for eroding the Board's independence and contradicts the provision that the Board shall hire its own staff.

Another issue relates to the Board's budget, which as a set-aside within the NAEP line item in the Department's Assessment, Statistics, Research and Improvement appropriation, is in fact the Secretary's responsibility to administer. The contradiction is painfully obvious—independence from the Department is incompatible with the Secretary's budget responsibility.

While I have highlighted a few of the areas of concern with respect to ambiguities in the role of the Board, attachment A contains a more complete analysis. However, I would point out that while we have not always agreed, for the most part, these issues have been handled openly and with a general spirit of cooperation with the Department.

At the moment, the Board and the Department are attempting to see whether, through a memorandum of understanding and/or delegations of authority, these ambiguities and contradictions can be addressed without the need for legislation. We are only at the starting point of this endeavor, so there is no way to be certain of the results. However, the working relationship between the Board and the Department has never been better, and we are hopeful at this time that legislation to resolve issues of the Board's independence will not be necessary.

Thank you very much; I would be pleased to answer any questions you may have.

Attachment A

POLICIES ADOPTED BY THE NATIONAL ASSESSMENT GOVERNING BOARD THAT WOULD REQUIRE LEGISLATIVE ACTION

JANUARY, 1991

BACKGROUND

During the past year, the National Assessment Governing Board adopted two comprehensive policy statements. In December 1989, the Board adopted positions on the Future of the National Assessment addressing nine specific policy issues. In No-

vember 1990, the Board adopted a statement on policy issues for the 1994-96 NAEP contract award addressing twenty-four specific policy issues. Some, but not all, of these policies cannot be implemented without amending the current NAEP authorization. This report has been prepared with the purpose of highlighting these policies.

In all cases, the policies that follow have been stated in a format that makes clear what is being recommended for legislation and may not be a word-for-word restatement of the original policy.

STATE-REPRESENTATIVE ASSESSMENTS

1. Provide a general authority to conduct State-representative assessments, with the frequency and subjects determined according to Board policy, with the full costs borne by the federal government, and with State participation voluntary.

RATIONALE

During 1990, both the assessment committee of the Council of Chief State School Officers and the National Forum on Educational Statistics have called for regular NAEP state-by-state assessment. The need for these data has been emphasized in an analysis of national education goals recently prepared by the Urban Institute. Also, State NAEP samples clearly are needed by the President and the Governors for their annual progress report on American education.

Without testing on a common instrument under common conditions it would be very difficult for citizens of any State to have sound comparative information on how much their students have learned. SAT's and ACT's are taken by self-selected groups. Commercial standardized tests have their "Lake Woebegone effect" in which most children appear to be above average. State criterion-referenced tests vary widely.

Congress authorized State NAEP on a trial basis in 8th grade math in 1990 and in 4th and 8th grade math plus 4th grade reading in 1992. No subsequent State testing is authorized. These trial State assessments are being closely monitored by the NAEP contractor and will be the subject of analysis across the country. An independent evaluation, required by law, is being conducted by the National Academy of Education; reports are due in late 1991 and 1993.

If these trial assessments are successful and the information useful, the support will likely be strong to proceed with State- based assessments in the NAEP program. If there are problems, changes in the assessments should be made. The underlying reason for State assessments is perhaps stronger today than ever and that is the need for sound, comparative data.

The Board believes that full federal funding is needed to help move to a testing program in which NAEP data are collected for every State, that there is a clear public interest in collecting such data, and that it is the only way a national progress report on American education can be complete.

Federal funding for many years has supported the collection and reporting of education data by State. These are essentially limited to "input" data including: enrollment, revenues, expenditures, program participation of students, pupil-teacher ratios, teacher characteristics, and high school graduation requirements. There is no dispute that a federal interest exists for the collection and reporting of such information at full cost to the government.

However, some analysts argue that there is only a very limited federal interest in collecting and reporting the outcomes of the application of these resources and requirements by State, that the outcome of the education process is a State concern and, therefore, State participation in NAEP should be limited to those voluntarily willing to assume part of the costs.

This view is perplexing. Almost no valid information exists about education outcomes at levels of analysis that bear on national and State policymaking. Policymakers at all levels are increasingly demanding such information and only the federal government is positioned, through NAEP, to provide it. Additionally, the amount of money involved is not excessive, particularly in comparison to other federal education initiatives.

The fact alone that President Bush and the Governors have, for the first time in our nation's history, set national education goals for the United States represents a sea change in the partnership in education between the federal and State governments. The fact that these goals describe outcomes underscores the concern for results. Responsibility for education is first of all a State matter; and on this basis States will be expected to do the "heavy lifting" in effecting reforms and improving education performance. It seems, therefore, fair and appropriate that the Federal

government, in fulfilling its role in the partnership, provide resources needed to help assess the results of State education improvement efforts.

Finally, while provisions related to voluntary costs to the States should be removed, those related to voluntary participation in the assessment should not. The intent is to encourage full State participation in NAEP by removing a federally wrought impediment; under no circumstance should participation by a State in NAEP be mandated, constitute a requirement for receiving any federal benefit, or in any way be other than voluntary.

ANNUAL NAEP ASSESSMENTS

2. Provide for annual rather than biennial assessments in the National Assessment of Educational Progress.

RATIONALE

It would be better to give NAEP on a rotating schedule of three subjects per year instead of maintaining the current practice of testing three to five subjects every two years.

Annual NAEP testing and reporting are necessary to provide timely and sufficient data for policy-makers and the public. This would allow NAEP to provide key measures of academic achievement for the annual progress reports on American education, called for last year at the Education Summit between President Bush and the nation's Governors.

For example, reading and mathematics plus one other subject could be tested in even-numbered years. Science and two other subjects could be tested in odd-numbered years. Annual schedules for data collection and reporting would produce cost savings by stabilizing work-flow and eliminating separate pretest samples for new items.

An annual NAEP would produce little net increase in test burden because the current every-other-year cycle includes up to five tests each testing year and extensive pretesting in off-years. Also, testing in three subjects annually would provide ample opportunity on a scheduled basis, perhaps every six years or so, to test in such subject areas as foreign languages, the arts, and economics, which now could only be assessed sporadically. The Board already is considering these subjects for possible assessments in 1996. It is important to test a range of subjects to discourage any narrowing of the curriculum.

NAEP conducted annual testing from 1969 until 1980 when testing was placed on a biennial schedule to save funds, so this proposal is not without precedent. Amending the current authority to permit annual testing and reporting rather than mandating a biennial cycle will provide more flexibility in the conduct of NAEP and allow it to be more responsive as information needs evolve. Of course, the number of subjects actually assessed would be subject to the availability of appropriations.

RESTORING LOCAL OPTIONS TO USE NAEP

3. Eliminate the prohibition against the use of NAEP test items and reporting below the State level by States, school districts, and non-public schools and school organizations. Modify the NAEP confidentiality provision to permit States, school districts, and non-public schools and school organizations to use NAEP for providing results on individual districts, schools and students. The cost of such uses of NAEP and of reporting would be paid by the requesting agency. Reporting of individual student scores would continue to be prohibited.

RATIONALE

These changes are related. They would restore the local option to use NAEP that States, localities, and schools previously had prior to 1988. The Governing Board has never suggested—nor does it now—that the organization which conducts NAEP should establish samples in individual schools and school districts as part of the regular NAEP contract and report the results. However, it is highly desirable that States, school districts, and schools have the option—if they wish to pay for it—of using NAEP to measure their own schools, as some have done in the past.

During the 1970's and 1980's a number of States used NAEP items and even NAEP tests to gather information on school and school district performance. NAEP tests and test items also had been used, apart from the regular assessment, for individual student testing in many State and local testing programs. All this activity was a desirable effort to provide useful data to students, parents, and schools.

However, in 1988 Congress enacted two provisions under NAEP that prevent State and local use of NAEP. The first prohibits any use below the State level of NAEP "test items and test data employed in the [State NAEP] pilot program." Although the statutory prohibition pertains only to questions used in the state-by-state assessments of 1990 and 1992, NCES has applied it to all NAEP exercises starting in 1990. The second provision requires the Commissioner to ensure "that all personally identifiable information about students, their educational performance, and their families and that information with respect to individual schools remain confidential . . ."

In July 1990, the National Forum on Educational Statistics, a group of 50 State representatives and federal agency officials convened by NCES, urged that States, "if they wish, [should be able to] analyze . . . student achievement [on NAEP] . . . so that comparisons could be made among education units by significant subgroups."

Even though supporters say the prohibition and the confidentiality provision protect local interests and family privacy, they also have an opposite effect—of denying the previously available option of using NAEP if localities wish to do so. The Board believes that these two provisions should be amended to clarify that they apply to the federal government and its contractors in the conduct of NAEP; to permit State and local use and augmentation of NAEP at local option and cost and pursuant to procedures established by the Commissioner to ensure test security, uniform administration and valid reporting; and to prohibit public reporting of individual student scores.

Large-scale use of NAEP raises serious issues of test security and uniform test administration and reporting. If NAEP tests are purchased by schools, given without supervision by thousands of classroom teachers, and reused for many years—as is done now with commercial tests—there is serious danger local results will be inflated and comparisons with national NAEP norms rendered invalid. Because of differences in motivation and possible preparation there would also be a serious problem in comparing the results of students who took NAEP knowing they would get individual scores with those in the national and State samples who are assured NAEP can't "count" for them. Also, the design of each NAEP subject area assessment—now divided into blocks for matrix sampling—would have to be modified to provide for individual student testing and reporting.

These concerns, however, can be addressed and should not be the basis for denying the use of NAEP and NAEP iten.s by States and local education entities.

INTERNATIONAL COMPARISONS

4. Provide authority for a regular international component for NAEP.

Rationale

NAEP should have a regular international component funded through its regular appropriation. The purpose of this component would not be assessment hegemony, i.e. transforming NAEP into the primary instrument by which participating nations compare their performance. Instead, the purpose of an international component in NAEP would be to provide information to help interpret and understand NAEP results, just as State-representative data are intended to enhance understanding of the performance of State education systems.

The President and the Governors have set national education goals in an international perspective (e.g. By the year 2000, United States students will be first in the world in mathematics and science) because a fuller understanding of the performance of U.S. students must take into account the education performance of our neighbors, trading partners and competitors.

Making valid international comparisons is a very complex undertaking. Curriculum and instructional practices, language and culture, student composition, provisions for special education and disadvantaged students, rate of high school completions, and the composition of in-school age groups all vary widely from country to country and prevent easy comparisons. However, if we are to have effective benchmarks for our own education practices and performance, they must be informed by expectations held for comparable students in other countries and the aggregate performance of those students.

Current law charges NAGB with "identifying appropriate achievement goals for each age and grade in each subject area tested under [NAEP]," "developing standards for interstate, regional and national comparisons," and "taking appropriate actions to improve the form and use of [NAEP]." Faithfully fulfilling these obligations demands not merely describing what U.S. students know and can do, but determining whether it is good enough. Making this determination should involve appropri-

ate analysis, in all subjects where practicable, of the performance of students in other nations.

For example, in setting achievement goals, it is essential to assure that they are both challenging and reasonable; examining the curricula of other countries would help in making this determination. Where other countries have comparable curricula in a particular subject area and a comparable student sample could be drawn, it would be useful to analyze respective patterns of student performance. It is not likely that this would involve numerous countries for each assessment, and it is not intended that such efforts would amount to a true international assessment. Information collected under the international component would be directed solely at informing U.S. results.

In 1988 an international comparison, using NAEP items in mathematics and science, was conducted by Educational Testing Service under grants from the National Science Foundation and the U.S. Department of Education. A similar comparative assessment, with similar funding, is planned by ETS in 1991. It will use items developed jointly by participating countries, including some from NAEP. Also planned are international assessments in science and mathematics in 1994 and 1998 to be conducted by the IEA, the timing of which coincides with the likely NAEP schedule for these subjects and which may be linked with NAEP. While the Board applauds these efforts, the fact is that they are conducted outside the regular NAEP framework. Thus, they are not a dependable source of information for developing policy (e.g., setting subject area achievement levels) and are not subject to NAGB policy.

With authority for an international component, NAEP could do a better and more complete job of reporting to the American people on the performance of American students. Having specific legislative authority for this purpose is essential to avoid reliance on data intended for other purposes, to assure orderly and systematic planning, preparation and data gathering, and to coordinate policy development with other agencies conducting international assessments.

ASSESS RANGE OF SKILLS

5. Modify the NAEP statement of purpose to clarify that assessment will cover the broad range of skills and knowledge (not just the basic skills) in the subject areas to be assessed.

RATIONALE

This clarifying amendment should be non-controversial. It reflects the view of subject matter and testing experts, expressed through the consensus process and in public hearings we have held, that NAEP assessments should include an examination of the higher order thinking skills of U.S. students as well as "basic skills." Our policy and practice already incorporate this view, but we believe specifying it in the law would be useful.

REGIONAL SAMPLES

6. Replace the requirement to employ regionally-representative samples and produce regionally-representative data with language that would make doing so a permissive activity.

RATIONALE

Regional reporting was from the start of NAEP intended as a substitute for State reporting. From a policy perspective, there is little cause for action that can be derived from such data, and there is little evidence that such reporting has had any effect on education policy.

With the possible advent of regular, predictable State level data collection, the need for drawing regional samples declines. While we do not advocate prohibiting regional samples and reporting, we do envision a time in the future when such sampling and reporting may be unnecessary. Thus, the Board should have the flexibility to recommend its application in the future consistent with other changes in NAEP.

INDEPENDENCE OF NAGB

7. Eliminate ambiguities and conflicting provisions in the NAEP legislation regarding the role and responsibilities of the National Assessment Governing Board (NAGB).

RATIONALE

The NAEP legislation gives NAGB the authority to "formulate the policy guidelines for the National Assessment," to "exercise its functions, powers and duties . . . independent of the Secretary and the other offices and officers of the Department of Education," to "exercise its independent judgment, free from inappropriate influences and special interests," and to "hire its own staff."

Taken together, these provisions could not be more clear in their intent to establish an independent policy body for NAEP insulated from the jurisdiction of the Department of Education. However, other provisions reduce this clarity.

Under the NAEP legislation, the Commissioner of Education Statistics carries out NAEP "[w]ith the advice of the National Assessment Governing Board (emphasis added]." Further, the law provides that NAGB "shall have the authorities authorized by the Federal Advisory Committee Act (FACA) and shall be subject to the open meeting provisions of that law." These provisions have been interpreted by the Department to mean that NAGB is an advisory committee subject to Department policies and procedures for advisory committees and that NAGB policy decisions must only be given consideration by the Secretary. Although the Department now agrees that NAGB is not an advisory committee, it continues to oversee NAGB activities as if it were, due in large part to the reference to FACA. Still unresolved is the degree to which the Department is obliged to observe policy established for NAEP by NAGB, rather than merely to consider it as advice.

Another issue relates to the NAGB budget, which, as a set-aside within the NAEP line item in the Department's Research, Statistics and Improvement appropriation, is in fact the Secretary's responsibility to administer. The contradiction is painfully obvious—independence from the Department is incompatible with the Secretary's budget responsibility. So far this issue has been managed in a spirit of cooperation, but it remains a potentially fatal flaw in NAEP's governance design.

There is a corollary to the budget issue. The Department has asserted that the Secretary's responsibility for administering and monitoring Department appropriations brings NAGB activities under its jurisdiction. Where compliance with government-wide requirements for fiscal accountability are at issue there is no disagreement, but assertions by the Department that its fiscal authority under the current arrangement extends to review and clearance of the intellectual activity of the Board and its staff in conducting its work is incompatible with NAGB's statutory mission and is simply unacceptable.

Another part of the law provides that "The Secretary may appoint at the direction of the Board" up to six technical employees under an excepted service appointing authority. Once again, while managed in a spirit of cooperation, this provision has the potential for eroding NAGB independence and contradicts the provision that the Board shall hire its own staff.

An area of ambiguity relates to evaluation of NAEP. Evaluation, particularly as prescribed under the NAEP legislation, is intended to bear on policy formulation. A commonly accepted standard for objectivity involves insulation of the evaluator from the subject of the evaluation. The NAEP legislation places responsibility for evaluation of NAEP with the Commissioner and is silent about the evaluation responsibility of NAGB. We believe that this is the opposite of sound practice for assuring objective evaluations and would advocate for an amendment that would add to NAGB's functions the formulation of policy for the conduct of evaluations of NAEP

While the law is very specific regarding the balance of responsibilities between the Secretary and the Board in making appointments to the Board, it is silent with respect to how NAGB will organize itself in performing its functions. The Board believes that the independence provisions in the law authorize total autonomy with respect to its internal workings. The Department believes this to be true with one exception—the appointment of the Board Chairman by the Secretary. Although this issue has been managed in a spirit of good will and cooperation, it remains a major point of contention.

These issues are fundamental. Their existence erodes the principles on which NAGB was established: independence from the Department of Education, tripartite check and balance NAEP governance structure, and freedom from inappropriate influences and special interests.

The goal that the Board seeks is no more than what Congress intended—a truly independent body whose policies direct the conduct of NAEP. The NAEP legislation does not define independence, but the Board believes that it includes three factors: (a) freedom from inappropriate influences and special interests in making administrative and policy decisions; (b) assurance that policy decisions made by NAGB will

be implemented in the conduct of NAEP; and (c) full, easy and timely access to information that NAGB needs from a variety of sources, including the Department, the Congress, education practitioners, and the NAEP contractor.

Legislation designed to achieve these ends could take many forms, from modifying or eliminating some of the provisions cited above to establishing NAGB with its current authorities as an independent agency outside of the Department. Should the Department of Education be contemplating changes in NAGB's authority as part of its reauthorization proposal, the Board would look forward to discussing this matter.

NAGB MEMBERSHIP

8. Provide for the continuing membership of the present Board according to current terms and for continuing the current provisions for filling vacancies. Add a provision that explicitly limits Board members to two consecutive terms of up to four years in each term. Encourage the Secretary to consider a candidate's previous experience on the Board and the overall experience of the Board when making appointments.

RATIONALE

The current legislation contains provisions for the transition in governance from the Assessment Policy Committee to NAGB. These provisions are now inapplicable and will be removed whenever NAEP is reauthorized; new provisions should leave no doubt about the continuation of the current membership.

During one thirteen month period in 1989-90, 14 of the 23 appointed members of the Board were replaced. This turnover rate is too high to assure the stability and continuity that was envisioned in providing for terms of up to four years. Providing for two terms and taking Board experience into account in making appointments will help assure stability.

Attachment B

POSITIONS ON THE FUTURE OF THE NATIONAL ASSESSMENT

APPROVED: December 9, 1989
At Meeting in Austin, Texas

The National Assessment Governing Board, mindful of its statutory responsibility to seek to improve the National Assessment of Educational Progress, hereby adopts these positions and recommendations:

1. *The National Assessment of Educational Progress should provide information for an annual report card by testing at least three subjects each year.*

The exact configuration should be determined after a careful analysis of cost and management considerations involved in such an increase over the current every-other-year testing cycle.

RATIONALE

Annual NAEP testing and reporting are necessary to provide timely and sufficient data for policy makers and the public. Under its current schedule, however, the National Assessment is woefully incomplete. Reporting assessment data each year will allow NAEP to become the key measure of academic achievement in the annual Report Cards on American schools, called for in September at the education summit between President Bush and the nation's governors. Also, regular annual data are needed from NAEP to replace the Education Department's annual "wall chart" which relies on SAT and ACT scores.

2. *NAEP should move as quickly as feasible to full state participation in all subjects and all three grade levels (4th, 8th, and 12th) tested. No state, however, should be compelled to participate. The federal government should pay the full cost of the state-by-state NAEP program.*

In 1990 each state must pay about $100,000 to take part in the state-by-state mathematics assessment. Several have cited this expense as a reason for not participating.

RATIONALE

The Governing Board wishes to have a full testing program in which NAEP data are collected for every state. There is a clear public interest in obtaining such data;

it is the only way a National Report Card can be complete. The Board believes it would be inadvisable to require states to participate in NAEP. However, full federal funding would make it less likely that states would decline to take part. By having NAEP testing in all states, the Board also wishes to end duplicate national samples as soon as feasible.

3. *The Governing Board urges Congress to remove the prohibition against the use of NAEP tests and data reporting below the state level.*

Publication of individual student scores should continue to be prohibited. Any additional costs for testing below the state-level samples should be paid by participating states or localities.

RATIONALE

The clause to be eliminated—Subsection (4)(C) of P.L. 100-297—reads as follows: "The use of National Assessment test items and test data employed in the pilot program authorized in subsection (2)(C) to rank, compare, or otherwise evaluate individual students, schools, or school districts is prohibited."

This change would permit NAEP test items and data to be used at the levels—school district and school building—where many important educational decisions are made. It would provide helpful information to parents and others interested in school improvement. By making NAEP more useful to local schools, offering this option would encourage them to participate in national and state sampling. States and localities would be able to tie their own regular testing programs into NAEP. Before the prohibition was enacted in 1988, many schools and districts requested NAEP reports on students who participated in national sampling; several states administered NAEP items.

4. *NAEP should establish international samples which would participate regularly in the Assessment program.*

Other nations should be invited to participate in NAEP on a regular schedule. NAEP tests should be administered to representative samples of their students, which might serve as reference points for achievement in American schools.

RATIONALE

The Report Card for American students is incomplete if it does not include a regular barometer of academic achievement in an international context. Today there is no such barometer.

5. *The Governing Board believes the release of NAEP data must be speeded up. It requests its staff to report by March on what steps could be taken to reduce the time for reporting NAEP test results, including possible design changes if needed.*

It has taken 21 to 24 months after testing to report NAEP results. This is scheduled to improve to 15 to 18 months for the 1990 assessment. But NAEP will still be much slower than the College Board and American College Testing Program summary reports and almost all state testing programs, which usually make data public in less than six months.

RATIONALE

The long delay in reporting severely reduces NAEP's usefulness and impact. It will be difficult for NAEP to serve as an annual Report Card if it is always more than a year late.

Money is not the cause of the lengthy delay. Rather, it stems from NAEP's highly complicated survey design and several basic decisions about how it is conducted. These issues should be studied carefully, including any trade-offs that might be necessary if NAEP design, test construction, and testing practices are changed. The study will include a cost-benefit analysis of any recommendations.

6. *The Governing Board believes the time needed to develop NAEP exams must be reduced. It requests its staff to report by March on what steps could be taken to shorten the time for planning and developing NAEP exams.*

Each new test now takes about 30 months to develop. This includes the consensus process for goals and objectives, preparation of test specifications and questions, review of questions by NAGB, and reviews of cognitive items and background questions by the Office of Management and Budget and the Education Department.

RATIONALE

The study should include all aspects of the test preparation cycle. Special examination should be made of the impact on NAEP of the Paperwork Reduction Act of 1980; reviews by OMB and the Education Department presently take a total of eight

months per cycle. The study will include a cost-benefit analysis of any recommendations.

7. *NAEP exams should be revised to include the full range of knowledge and skills from basic skills to advanced subject-matter knowledge and analytical, integrative skills.*

Test planning committees and item writers should include a significant number of test objectives and questions on NAEP exams that assess higher-level analytical skills and knowledge as well as basic skills.

RATIONALE

Attention to higher-level skills and knowledge as well as basic skills is clearly in line with efforts to strengthen the school curriculum. It will permit NAGB to establish high standards when it sets goals for NAEP tests. The NAEP math exam has already been changed in this direction.

To achieve this goal NAEP exams may have to include more open-ended questions and essays instead of relying as heavily as they do on multiple-choice items. The multiple-choice questions themselves may also be changed to test more higher-level skills and knowledge than they do now.

8. *NAEP should obtain data on additional important groups.*

Sample sizes should be enlarged to provide information on groups such as low-income children and those attending private schools.

RATIONALE

This would provide important information for educators and policy makers and for analysis of schools. The definition of low-income students may be difficult, but NAEP could use those who qualify for participation in federal free or reduced-priced lunch programs or children who attend Chapter I schools.

9. *Change in the governance structure of NAEP should be considered by Congress.*

The role and responsibilities of the Governing Board should be clarified by new legislation.

RATIONALE

The current system of divided authority may lead to confusion and disputes which would hamper the NAEP program. Governance and administrative supervision of the National Assessment should be rationalized and focused. Overlap and duplication of functions of NAGB and NCES staffs should be eliminated.

Senator BINGAMAN. Dr. Stewart, go right ahead.

Mr. STEWART. Thank you very much, Senator Bingaman and Senator Pell.

I am Don Stewart, president of The College Board, which is a nonprofit association of close to 3,000 schools and colleges. Perhaps I should also mention that I am president emeritus of Spellman College in Atlanta, GA and a graduate of Grinnell College a few years before David Evans left.

In recent weeks I have been meeting with Secretary Alexander to offer the resources of The College Board in helping to realize the vision of "America 2000" as well as with Governor Romer to help develop the indicators that will go into the first National Report Card that will be issued in the fall. I very much hope that The College Board can be a resource and of assistance to this subcommittee as you deliberate on how to improve schooling in America.

The College Board is prepared to contribute in any way possible to the emerging national assessment and school improvement agenda, and in my brief remarks this morning I would like to try to link those objectives with your considerations for OERI.

In my view, we will succeed in reaching our national education goals if 1) there is a consensus on standards and what students are expected to learn—the "what" of schooling; and 2) if teachers and schools are empowered to support all students in achieving these

high standards—the "how" of schooling; and 3) if there are appropriate means to measure both in a formative and a summative manner educational progress—the "how well" of schooling.

I urge the Office of Educational Research and Improvement, OERI, to foster research and analysis in each of these areas and that it be given sufficient resources to do so in depth. We need thorough research on current forms of assessment, on the critical distinction between diagnostic testing and testing for accountability, and on the implications of current national testing proposals for all students, the schools they attend, and the teachers who instruct them.

Let me deal briefly with the "what", the "how" and the "how well" of education in somewhat greater detail.

The "what" relates to standard setting, and for the past 91 years, The College Board has brought together educators from the secondary and collegiate levels to reach agreement on educational objectives and test content across a wide variety of subjects. These experiences resulted in a publication, "American Academic Preparation for College", which describes what students ought to know and be able to do upon graduating from high school. The so-called "green book", a copy of which I will submit for the hearing today, describes learning outcomes in English, the arts, mathematics, science, social studies and foreign languages.

And you might note, Mr. Chairman and members of the subcommittee, that The College Board's list of core subjects includes the arts and foreign languages, which are missing from or at least not stressed in "America 2000".

The mathematics community has led the other disciplines in building consensus on standards. Standards recently issued by the National Council of Teachers of Mathematics, NCTM, are groundbreaking. The next step is to move to a level of detail that is sufficient for both course construction and the assessment of learning outcomes. The College Board is working with NCTM and the Mathematical Association of America to develop such elaborated standards at the 8th and 12th grade levels.

All of these efforts at standard setting are powerful because educators and educational associations are involved. Standards in turn are translated into learning through the teaching process and teacher empowerment that relates to the "how" of education.

Unfortunately, the current discussions of national assessment have paid too little attention to the critical role of teachers in implementing standards, and there can be no success, in our opinion, without the involvement of teachers. New standards will require teachers to improve current skills and develop new ones, to invent new methods and to experiment. This will not happen accidentally or indirectly. The College Board is planning to work with NCTM and MAA to undertake strong professional development in support of high school math faculties.

Finally, assessment is required to determine how well our educational system is doing. The recommendations of the National Education Goals Panel for national anchor examinations, the new American achievement tests, and the proposal for the development of one or more comprehensive tests of academic excellence con-

tained in S. 1016 all represent a challenging vision for the future of assessment in our country.

I do not, however, support the idea of a single national achievement test. Multiple learning outcomes over time require a diversity of assessment practices. Moreover, I am skeptical about the practicality of having one test that effectively serves the needs of accountability on the one hand and student development, i.e., learning, on the other.

Much of what the national assessment movement is hoping to accomplish is already embodied in The College Board's advanced placement program, which Secretary Alexander has described as the best test we have today for major skills. AP is a curriculum-based test with teachers involved in the test development, in the syllabus development, and in the testing.

Having been given the red light, I will skip to the end of my statement, and I would like to say in passing that we feel equity must also be an important consideration in your deliberations. High standards without equity are destructive and morally unfair. Students must be prepared to meet higher standards. And we would like to participate in that process as well.

Mr. Chairman, members of the committee, I appreciate this opportunity to appear before you, and I apologize for talking so long.

Senator PELL. [Presiding.] Thank you very much indeed, and your full statement will be included in the record as if read.

[The prepared statement of Mr. Stewart follows:]

PREPARED STATEMENT OF MR. STEWART

Senator Pell and members of the Subcommittee on Education, I am Donald M. Stewart, President of the College Board. I very much appreciate the opportunity to comment upon testing and assessment issues as they relate to the reauthorization of the Office of Educational Research and Improvement (OERI).

Founded in 1900, the College Board is a national nonprofit membership association of more than 2,800 schools and agencies in higher and secondary education committed to promoting educational opportunity and improving academic standards. The Board sponsors programs in guidance, admissions, placement, assessment, financial aid, and credit by examination to assist the school-to-college transition of some four million students each year. Throughout our history one of our activities has been to establish standards through collaborative efforts involving school and college faculty. The Board also sponsors research and provides forums to discuss common problems in education.

As you know, this is a time of extraordinary opportunity for education thanks to the leadership and vision of President Bush and Secretary Alexander, Congressional leaders like the members of this Subcommittee, and the nation's governors. For the first time in our history we have a set of national education goals. Achieving these goals can improve the quality of American education not just for some, but for all. And achieving these goals is imperative since the stakes are high and the end results will determine our educational future and even our status as a democratic nation.

Others have provided testimony on specific issues and agendas for OERI and on expansion of the National Assessment of Educational Progress (NAEP). My observations today will address the broad challenge facing the country in advancing the national education goals and designing assessment instruments to measure progress towards them.

In recent weeks I have met with Secretary Alexander to offer the assistance of the College Board in realizing the vision of *AMERICA 2000*. I also have met with Colorado Governor Roy Romer, chair of the National Education Goals Panel (NEGP), to help in developing the indicators that will go into the first national report card to be issued this fall. I make the same offer of assistance to the members of this subcommittee as you deliberate on how to improve schooling in Amer-

ica. The College Board is prepared to contribute in any way possible to the emerging national assessment agenda.

In my view we will succeed in reaching our national education goals if:

—there is a consensus on standards and what students are expected to learn—the "what" of schooling;
—teachers and schools are empowered to support all students in achieving these high standards—the "how" of schooling; and
—there are appropriate means to measure, both in a formative and summative manner. educational progress—the "how well" of schooling.

Furthermore, we must take steps along the way to ensure that all students are expected to meet high standards and have access to quality education. High standards without equity are neither productive nor fair. Instead they are destructive and morally unfair.

I urge the OERI to incorporate research on each of these areas—WHAT all students are expected to learn, HOW teachers and schools support them in learning and HOW TO MEASURE their progress—in its upcoming agenda. In particular, I urge 0ERI to conduct thorough research on the various forms of assessment currently being discussed, to distinguish between assessment for accountability and for learning, and to carefully examine the implications of these proposals for all American students, for the schools they attend and for the teachers who instruct them.

Let me elaborate.

ESTABLISHING STANDARDS. THE "WHAT" OF SCHOOLING

We must be sure to integrate assessment and curriculum without blurring the distinction between the two. By curriculum, I mean the objectives of education, or the "what" of schooling. By assessment, I mean the gathering of information about the extent to which the objectives have been achieved—the "how well" of schooling. Agreement upon the "what" of schooling must precede agreement upon ways to measure the "how well" of schooling.

In that regard, I am pleased that both the NEGP and America 2000 acknowledge the importance of first establishing standards. The NEGP endorses the creation of a framework reflecting "what the nation wants young people to know and be able to do as a result of their years in school" as the first sequential step in the development of their proposed examination system. The America 2000 strategy calls for the development of "new world standards" and a new nationwide examination system tied to these standards.

The College Board has a wealth of experience in bringing together educators from the secondary and collegiate level to reach agreement on educational objectives and test content across a wide variety of academic subjects. These experiences resulted in a publication—*Academic Preparation for College* (also known as the "green book")—describing "what students should know and be able to do" upon graduating from high school.

The "green book" describes learning outcomes in English, the arts, mathematics, science, social studies and foreign languages. It also identifies and describes basic academic competencies—reading, writing, speaking, and listening. A subsequent set of College Board publications, the "rainbow" series, provides specific curriculum and instructional suggestions about how to achieve the results outlined in the "green book."

The experiences of the mathematics community in establishing standards is also illustrative of what must be involved in reaching consensus in other disciplines. Standards recently issued by the National Council of Teachers of Mathematics (NCTM) are an important and first step in that direction. In fact, the math community was greatly assisted in its consensus-building activities by the College Board's publication of *Academic Preparation for Mathematics*. The next step is to move these expressions of values to a level of detail that is sufficient for both course construction and the assessment of learning outcomes. The College Board is currently working cooperatively with the NCTM and the *Mathematical Association of America* (MAA) to develop such "elaborated" standards at the 8th and 12th grade levels. Just as the NCTM standards serve as an example of a first step, so can the College Board/NCTM/MAA collaboration serve as an example of the requisite second step.

Both of these efforts are powerful because educators and educational associations are involved. This is as it should be. I urge OERI to build upon the standard-setting experiences already underway at the College Board and in the math community.

EMPOWERING TEACHERS. THE "HOW" OF SCHOOLING

Standards are translated into learning through the teaching process. Unfortunately the current discussions of national assessment have paid too little attention to the critical role of teachers and other educators in implementing standards and there can be no success without this involvement. The assumption is that teachers should and will respond to the national education goals, but little is being done to include them directly in the process or to support them in their classroom efforts. This must change or the goals will not be reached.

New standards will require teachers to improve current skills and develop new ones, to invent new methods and to experiment. This will not happen accidentally or indirectly. It must be planned and executed. Professional development and training efforts must be put in place to involve teachers in the setting of standards and to assist them in preparing students with a full range of abilities to meet outcome expectations.

At the College Board we are planning to work with NCTM and MAA to undertake a strong professional development effort in support of high school math faculty. This effort will focus on the central role that teachers should play in linking standards and assessment.

I urge OERI to consider the expansion of these types of efforts to other disciplines and to more and more teachers.

CHOOSING ASSESSMENT INSTRUMENTS. THE "HOW WELL" OF SCHOOLING

I have been describing an education process that starts with the establishment of standards, involves teachers in the translation of these standards into classroom and other learning activities, and culminates with the assessment of learning outcomes.

This final phase of the process leads to the question of whether or not there should be a national examination (or system of examinations). It also raises the question of what kinds of testing should be utilized and developed. Many suggest that we move away from traditional multiple-choice examinations to more diagnostic forms of assessment that may do a better job of both measuring and encouraging educational achievement. And finally the issue of testing for accountability vs. testing for individual learning and development must be addressed.

The recommendations of the NEGP for national anchor examinations, the proposals contained in America 2000 for the new (American Achievement Tests), and the proposal for the development of one or more comprehensive "tests of academic excellence" contained in S. 1016 all represent a challenging vision for the future of assessment in our country. The College Board can contribute a great deal to realizing this vision and is already doing so.

I do not, however, support the idea of a single national achievement test; multiple learning outcomes over time require a diversity of assessment practices. For this reason, the College Board is looking closely at how to develop diversified assessments and how to implement an examination system consistent in each of its parts and coherent in its whole.

Last fall the College Board announced revisions to the SAT designed to make the test even more responsive to the changing educational needs of the nation and a new generation of students. The redesigned SAT, to be administered in the Spring of 1994, will consist of two parts. SAT-I (Reasoning Tests) is a measure of broad verbal and mathematical reasoning skills that students learn through their courses and outside of school as well. SAT-II (Subject Tests), more directly reflective of the high school curriculum, will be enhanced and expanded to include a new writing test that combines multiple-choice questions and a direct writing sample, new tests in Asian languages and proficiency in English as a second language. These changes will benefit schools (by providing more information for guidance and counseling purposes and curriculum evaluations), colleges (by offering improved information for admissions, advising and placement) and students (by aligning the tests more closely to current classroom practices and by offering a variety of tests to demonstrate individual strengths).

Much of what the national assessment movement is hoping to accomplish is already encapsulated in the College Board's Advanced Placement (AP) program described by Secretary Alexander as "the best test we have today to honor proficiency in major skills." In fact, President Bush is proposing to use performance on AP examinations as the basis for the Presidential Citations for Educational Excellence and AP data will be used as an indicator on the 1991 Report Card issued by the National Education Goals Panel.

The AP program has grown dramatically over the last decade. Since 1985, the number of minority students participating in the AP program has increased consistently, as has the number of AP secondary schools with predominantly minority enrollments. Minority participation in AP (24 percent) is about the same as the minority percentage of the secondary school population.

The current AP program provides students with the opportunity to complete college-level studies while in secondary school. It is curriculum-based with teachers directly involved in the development of courses and of examinations designed to measure achievement utilizing both multiple-choice and student-produced answers as forms of assessment. Multiple-choice questions are included to sample broadly from the content of the course, while the free-response format enables a student to demonstrate the ability to "do" some valued aspect of the subject.

In addition to these assessment tools, the College Board, in conjunction with the Educational Testing Service (ETS), has begun to plan how we might further address the current calls for a multiplicity of assessment forms. As we look to the future, I think it will be possible to create curriculum-based, AP-like tests that incorporate performance-based components and are used generally in high schools for all students; relate high school curriculum to outcomes; and involve teachers in reading and grading on a decentralized basis.

While our ideas are still in the formative stage, we are convinced that an approach and process is needed that links (1) course development by appropriate academic disciplines working together in a time honored collegial manner; (2) identification and conduct of important and needed teacher preparation and support; and (3) development of assessment approaches that measure mastery of desirable learning outcomes.

Finally, much of the current discussion of national assessment has tended to blur the distinction between testing for accountability with testing for individual development (learning). In practice there is frequently a difference between the two and this must be understood by all those concerned with improving education.

Testing for accountability involves the gathering of information about the performance of groups of students in order to inform educational policymakers including government officials—about the effectiveness of schools at the state, city, district and school level.

Testing for individual development provides information to teachers and students that will help guide instruction and learning in the classroom; it provides feedback to students, parents and teachers so that instruction can be shaped to help overcome weaknesses in learning. For example, the SAT has limited value as a test of accountability, and it is far more appropriately used to assess individual learning.

Tests for individual development differ from tests utilized for accountability in two important ways. A test used for accountability must cover a broad range of content representing a good cross section of valued educational objectives. In contrast, a test used for individual development is narrower in scope but deeper in content; and a test must provide information to both the teacher and the students about the mastery of very specific subject matter by individual learners. Furthermore, tests for learning must be given at the individual student level. This is not true for tests of accountability; carefully selected samples can provide sufficiently reliable data for this purpose.

I urge OERI to pursue an agenda that will clarify the purposes of testing and distinguish between testing for accountability and testing for individual development. Though it is technically possible to do so, I remain skeptical about the practicality of developing one test that will serve effectively both the purposes of accountability and individual development.

ENSURING HIGH STANDARDS FOR ALL STUDENTS

Before closing, I want to underscore the need to guarantee access and equity for all students. At a time in our history when cultural, racial and ethnic diversity has never been greater, it is essential, as a matter of economics as well as social justice, to assure quality education standards for everyone. We must not allow the rush towards the establishment of standards and the creation of new assessment systems to obscure (or worsen) the extreme inequalities of opportunity that currently exist for minority and disadvantaged students. We must monitor every step, every plan, and every action so that equity is advanced while high standards are achieved.

If you will allow me, Mr. Chairman, I would like to mention in this context a national campaign launched by the College Board, Equity 2000, which was announced here in Washington on June 4. Designed to raise the college-going rates of minority and urban students by the year 2000, the program is based on research that statistically demonstrates the importance of student mastery of algebra and geometry as a

prerequisite to pursuing and succeeding in higher education. The setting of standards, the preparation of teachers to meet those standards, and the utilization of assessment techniques to identify strengths and weaknesses in student progress are all an integral part of our Equity 2000 project.

In its first stage, Equity 2000 currently involves six local communities across the country. Senator Pell, I appreciate your attendance at our luncheon last week and note your pleasure in having Providence, Rhode Island as one of the sites. During 1991-92, Providence schools will offer all eighth grade students pre-algebra; in subsequent years these same students will be required to take algebra and then geometry. Students will be provided with intensified and on-going guidance and counseling to motivate and support them to seek a college education and to succeed when they get there. With support from a number of private foundations, we hope to establish a model that can be replicated and will lead to widespread and long term improvement in education for all students.

We cannot allow comparisons with the educational achievement of other countries, and their testing programs, to obscure the real strength of American education. We must remember no other country has an educational establishment which aspires to the degree of student participation, the pluralism of means, or the multiple layers of opportunity which exist in the United States. The achievements of other systems come from the fact that they have identified clear educational objectives toward which instruction for some students is aimed. The challenge for America is to establish objectives which are no less rigorous, but also to create the means by which they can be pursued successfully by all students.

CONCLUSION

I very much appreciate the opportunity to appear before your Subcommittee and to share my perspectives with you. I hope that my remarks on the importance of setting standards, the preparation of teachers to help students meet those standards, the use of specific assessment instruments and the importance of ensuring equity will be helpful as you continue your deliberations on the reauthorization of the Office of Educational Research and Improvement.

I close with a fervent hope that the commitment now being focused on educational reform will bring positive results for students, their parents and schools. Through dialogue, consensus and cooperation American education can surely emerge strengthened and revitalized.

Thank you.

Senator PELL. Dr. Stewart, you established standards within English, the arts, math, science, social studies and foreign languages. What was the process by which you developed those standards, and how did you choose those particular subjects?

Mr. STEWART. Mr. Chairman, the answer to your first question is over a long period The College Board in association with teachers and college faculty in those disciplines worked on the development of those standards. They are described in the "green book" and in individual books by subject.

We have what we call the "rainbow series" which breaks the fields down into those that you have just mentioned.

To answer your second question, there was a consensus that these were the essential fields of knowledge that provide both the skill and competency underpinnings and the subject matter knowledge needed to perform well in college, and frankly, in life.

So it was based on a perhaps somewhat traditional view of the nature of learning and knowledge and of a liberal arts curriculum.

Senator PELL. Thank you very much.

Senator Kassebaum.

Senator KASSEBAUM. Thank you, Mr. Chairman.

First, my apologies for missing the testimony; I have been trying to catch up.

Regarding testing, Dr. Stewart, what would be your suggestions, if you think a national test has some problems, in terms of being

able to work with collaborative efforts? As the Secretary was implying, perhaps each State and/or each district could adapt certain areas and focus that is important to that area.

Does that have any merit?

Mr. STEWART. It may certainly have merit, Senator Kassebaum. I would hope to see how that might work out. What I am hoping does not happen is that we lose the power of a national accountability test such as NAEP—and even if it goes down to district and school levels, that we still do sampling and aggregate data so that we understand the important trends and can make the comparisons among districts, State, or whatever, without going to the individual student.

I would hope that any tests that are focused on or used by whatever agency or entity for individual student testing would be ones that enhance learning, that teachers would be involved in developing and would not feel threatened by an externally-imposed test, but would in fact use testing creatively for diagnostic purposes, etc.

Also, tests tend not to follow political units, and knowledge is not so organized. I think there are many forms of national tests now, some that are no longer considered good because they do use multiple choice such as our own SAT, and standardized testing is something that is being questioned. But tests cut across multiple layers—government schools, whatever—and for the integrity of testing, I would hope that the units would not be political defined.

Senator KASSEBAUM. Maybe both of you could answer this. How well are teachers prepared to use the newer forms of assessment,such as performance-based assessment; how much retraining, if any, would be required; and how many of the teacher education programs, as a matter of fact, in the country are teaching prospective teachers about alternative assessments?

Mr. BOEHLJE. I'm not sure I am qualified to answer that from the perspective of NAEP because the type of assessment that NAEP is is a random sampling that is designed to evolve a reading of the population in general and not get to specific assessment. So it is very, very strictly limited, and it is a very specialized type of test.

Senator KASSEBAUM. Kansas did not participate in the last test that was done of 8th graders in math. I had some teachers tell me they were disappointed that Kansas had decided not to participate. I think that is a decision that has to come from the State, but obviously teachers, particularly those teaching in the subject, had some strong feelings. I don't know how the teachers at that point become a part of this process.

Mr. STEWART. Senator, if I may, my sense is that a great deal of groundbreaking work still needs to be done in schools of education and in service training programs and so on for the teaching profession.

At the same time, I think there is a groundswell in our country with a number of experiments going on in school districts, with teachers being helped to better understand the uses of testing for informing instruction, for looking for performance measures as opposed to just multiple choice. I think now there is a growing feeling in the country that performance-based testing is doable and important.

We have just restructured the SAT to make it far more performance-based in anticipation of this, and new technologies make it possible; much more testing will be taking place on computers, and I think as part of pedagogy, new approaches to testing and measurement are very much gaining momentum. But we have a long way to go.

Senator KASSEBAUM. Of course, we always wonder about the child who may fall through the cracks. That is one of my real concerns with testing early on, that then someone doesn't become identified. That's why I think, particularly in the early years, we have to be sensitive to how we approach testing.

Regarding NAEP, how much more would it cost to do it yearly rather than every 2 years, because I think the overall assessment has much merit. I think it creates an opportunity, without getting too specific, to show where a State stands. I didn't know if you had given any thought to the merits of that.

Mr. BOEHLJE. The discussion of cost has taken place. I am not privy at this point to the estimates on what the additional costs will be. I know the NAGB staff at one point was making a determination of that, and I can submit that to you at a later point.

Senator KASSEBAUM. But you do think it would be a good idea.

Mr. BOEHLJE. Yes, very definitely.

Senator KASSEBAUM. How would you go about encouraging a State like Kansas, for instance, that has chosen not to participate to participate?

Mr. BOEHLJE. That's a tough question.

Mr. STEWART. You might speak to its Senator. [Laughter.]

Mr. BOEHLJE. One of the important factors of this assessment process is that it is voluntary and that States like Kansas can elect not to be involved. I think that the reaction we are getting from this first reporting will be positive enough that most of the States are going to want to come on line. We are seeing the numbers increase already.

Senator KASSEBAUM. I would guess that's very true. With the attention and the publicity paid to this, which is really in many ways the first time we have had such a national focus on tests and education, it seems to me that States will want to participate. Pressure will come from those who believe it would be serious to be left out; that failing to participate, in and of itself, casts aspersions on the quality.

Mr. BOEHLJE. I guess the encouraging trend that I have seen so far—and it really has been just a very short time since this math assessment has been released—but when you look at the analysis that has been made—most of the press releases and statements that we're getting are from State departments of education—but they have not keyed on comparing themselves with another State, as a rule.

Senator KASSEBAUM. That's true.

Mr. BOEHLJE. They have looked at the assessment as a basic tool and said, "This shows that we have significant problems, and this shows some of the factors that appear to relate to these problems, and these are the areas that we have to address."

I think if that attitude is going on by the people who are currently involved in it, it should not be a deterrent to the remainder of the States hopefully coming on line.

Senator KASSEBAUM. I have been very impressed, and I think it has served a useful function.

Thank you very much.

Senator PELL. Thank you very much, Senator Kassebaum.

Senator Bingaman.

Senator BINGAMAN. Thank you, Mr. Chairman.

It seems to me logical that at some stage here, all of this would come together into one test. I mean, at some point, if you are going to test a 4th grader on how much mathematics that 4th grader knows, it is hard for me to understand how you explain to him that we need to give him one test to talk to his folks about how he is doing, and we need to give him another test to do the NAEP assessment nationally.

Do you agree that at some stage, it is all the same thing—I mean, that at some stage we set standards, we prepare a test that we all agree measures to those standards, and that is it?

Mr. BOEHLJE. Well, at some point it appears that it would be in everyone's best interest to have those items brought together. One of the things the NAGB board has discussed consistently is this whole question of testing and how do we make the test effective, how do we make students want to take the test when there is no individual reporting back, how do we address the fact that students may have 15 other tests that they have to take——

Senator BINGAMAN. So you are making the case that I am making, that it all ought to come down to one test in each subject, at whatever grades we decide to test.

Mr. BOEHLJE. That's an issue that has been addressed. One of the big problems that we see as a board and that the board has not come to a conclusion on is that the NAEP assessment is a very different type of assessment from a national test for individuals, and bridging that type——

Senator BINGAMAN. It is very different in what sense?

Mr. BOEHLJE. In the reporting sense.

Senator BINGAMAN. Oh, in the reporting sense, I understand. But you are still trying to get to the same core issue, which is how well is this 4th grader performing in math.

Mr. BOEHLJE. That's correct.

Senator BINGAMAN. And how well is this 4th grader performing in math relative to how he or she should be performing.

Mr. BOEHLJE. Yes.

Senator BINGAMAN. So that to that extent, I understand that you do yours at random, and you are trying to assess national trends and that sort of thing—but I guess I'm just saying that the logic drives you to a conclusion that if we come up with a national test, NAEP ought to administer it, our NAEP ought to go out of business.

Mr. BOEHLJE. It depends on what you want from NAEP. Yes, you could make that statement. If you want NAEP to continue to simply report trends and be independent from the national individual testing philosophy, then fine. The credibility of NAEP would indicate that it would lend itself very nicely to becoming the na-

tional test. But as I say, that is a very, very big step, and making the jump to becoming that test is very costly, very——

Senator BINGAMAN. Well, the Secretary said that in 1994, his plan and the President's plan is that NAEP will do State by State comparison testing—not of all students, obviously, but enough testing in each State that you can do State by State, as I understood what he said——

Mr. BOEHLJE. Yes.

Senator BINGAMAN [continuing]. In five subjects, at three different grade levels, 4th, 8th and 12th. I understand that you don't have the dollar figures as to what that would cost, but can you tell me if that is a realistic goal? Can you gear up to do that?

Mr. BOEHLJE. The time frame?

Senator BINGAMAN. Yes—in order that that test be administered in 1994.

Mr. BOEHLJE. I think it is possible on a test basis—and by "test" I mean on a specific, limited basis. I don't think it is possible on a national basis to gear the whole process up that fast.

Senator BINGAMAN. Well, now, you can't do it on too limited a basis and still have the State by State testing that is now being talked about, can you?

Mr. BOEHLJE. That's right.

Senator BINGAMAN. So I guess my question is can you do what is now being set out as the goal?

Mr. BOEHLJE. I suspect it can be done if the dollars are allocated to set the framework up to let the contracts for the test, but it is a tight time framework.

Senator BINGAMAN. And you are going to get back to us on the amount that will cost?

Mr. BOEHLJE. Yes.

Senator BINGAMAN. That would be very useful. Thank you.

Thank you, Mr. Chairman.

Senator PELL. Thank you very much, Senator Bingaman.

Thank you, Dr. Stewart and Mr. Boehlje, for being with us.

[Additional statements and materials submitted for the record follows:]

PREPARED STATEMENT OF MICHAEL A. RESNICK, ASSOCIATE EXECUTIVE DIRECTOR, THE NATIONAL ACADEMY OF EDUCATION

RESEARCH AND THE RENEWAL OF EDUCATION

SUMMARY OF FINDINGS

As Americans strive to reform and restructure schools and design programs that prepare young people for a new century, education policy and practice must be guided by the best that is known about education and all areas of inquiry related to learning, schooling, and young people.

According to a new report from the National Academy of Education, an honorary society of 75 of the nation's most distinguished researchers and educational leaders, current efforts to implement broad-based school reforms without adequate research to guide the direction of change will lead to failure. "Pushing for change without continuing to deepen our understanding of what we are doing will intensify the problems we seek to solve," says the study, *Research and the Renewal of Education*, which will be released next month.

The Academy's study asserts that if we want well-designed institutions, good leadership, high standards, and public support for education, we need imaginative re-

search that anticipates the future, takes on the biggest challenges, and is funded to support the most promising areas where breakthroughs are likely to occur.

The Academy suggests how the organization and character of research can be changed to improve its application to policy and practice. The report cites five notify areas expected to spark positive changes in schools. These include:

—*Assessment.* As educators debate the need for new forms and "applications of testing, a major investment is needed to develop new educational assessments. Research should improve the instructional relevance of testing, probe the social contexts of learning, and foster a rich view of thinking and creativity.

—*Active Learning over the Lifespan.* Research must be designed to understand how students can be intellectually engaged and encouraged to solve challenging problems as individuals and as a group—the kind of learning they will use in the real world. Research must focus on ways to help students take initiative, construct meaning for themselves, and develop thinking skills in new and unfamiliar settings. The Academy underscores the fact that learning exists outside of schools and that research must embrace education in its broadest contexts—including learning that takes place within families, communities, and in other settings.

—*Bolstering Achievement of Historically Underserved, "Minority," and Impoverished Groups.* More research is needed on the social backgrounds, cultural contexts, and learning dispositions of these groups as related to education, and on the institutional arrangements that prove most effective in expanding educational opportunities for such groups.

—*School Organization.* We need further inquiry into the social organization of schooling and the inner workings of schools as institutions. More research is needed to clarify how effective learning is organized, whether within school or beyond. The Academy notes that the structure of schools constrains the dissemination of new innovations based on research and that new organizational structures hold promise for more widespread use of research.

—*Connection to Teaching.* Connecting theory to practice is more than examining instruction effectiveness or devising new forms of professional development. It also means placing research in the service of teaching and school improvement. Teachers and researchers must be collaborators in constant communication with each other, but this will involve new roles for teachers and students, according to the Academy.

This new agenda, the Academy asserts, will create new support for research and its applications and lead to real results in schools. The Academy's goal is to develop research that will be useful in the future. "As in medicine, researchers must discover tomorrow's cures, not yesterday's leeches," the report says.

The study, funded by the Carnegie Corporation of New York, was directed by Michael Kirst of Stanford University and Diane Ravitch of Columbia University's Teachers' College. The executive director and principal writer of the study was Thomas James of Brown University.

The study argues that while researchers have not developed the equivalent of antibiotics or hybrid corn, investigations conducted in universities, laboratories, and in schools themselves have played a crucial role in shaping the structure and content of schooling and in deepening our understanding of the education process. Education research has led to fundamental reorientations in our knowledge of human development, learning theory and its applications, testing and assessment, the nature of disabilities, and curriculum design. " Education research has led schools to turn 'good' practice into 'best' practice," according to Michael Kirst.

Today, researchers are bringing new understandings in areas such as the process of reading, the craft of writing, the growth of reasoning skills, and the nature of cultural differences in learning that are leading to significant improvements in education. In addition, research is helping to shape public policy, such as in the design and implementation of mandates for equitably financing school systems in the states. (See attached sheet for examples of how research has influenced practice.)

The Academy argues that the answers to our education problems lie in creating new connections between what is discovered through research and what schools and teachers do in their policies, procedures, and instructional practice.

According to the Academy, educational problems cannot be solved by developing treatments that focus on one level of the system, such as state governments, school districts, or classrooms, or on just one aspect of the process under way in learning institutions, such as reading or mathematics.

Problems in Research

But the promise of research in shaping educational change, the Academy notes, is limited by constraints within and outside the research community. The research base is underfunded, limited in focus, and traditionally lacks connection to what happens in classrooms. Research studies tend to be small-scale, short term, and conducted in isolation.

Educational research as a whole is fragmented and theoretically diffuse, marked by a profusion of studies that too often lead down divergent paths to endlessly debated viewpoints and assertions. The paucity of longitudinal studies has resulted in an over-abundance of "snapshots," studies of specific treatments and interventions without a systematic knowledge base established over time and under varying circumstances.

Institutional research does not adequately take into account outside forces impinging on the educational settings under study, such as the social, cultural, and economic forces influencing dropouts, testing, and tracking.

In addition, the Academy assets, the whole enterprise is slow in responding to powerful new currents of fundamental research in disciplines touching upon the study of education.

Part of the problem with research is its funding base. Patterns of support for educational research are episodic and hampered by changing demands, vacillating leadership, unstable commitments, and institutional pressures. Most research on education is not funded at levels sufficient to allow intensive experimentation and collaboration with educational practitioners. Neither the federal nor the state governments fund leading centers of research well enough to give them status as centers of excellence and to communicate effectively with educational practitioners.

Funding for research through the National Institute of Education—for many years the largest source of research support—has been decimated since 1973, the peak funding year. Between 1973 and 1986 funding dropped nearly 80 percent before being adjusted for inflation, and there has been no significant increase since that time.

The U.S. Department of Education spends less than $120 million for research. In comparison with fields such as defense, health care, energy, and agriculture, the nation spends little on disciplined inquiry and research-driven experimentation for improving education.

Only $2 million of the 1992 budget request for the Office of Educational Research and Improvement covers field initiated research, and even this small amount represents a 100 percent increase over last year. Consequently, the entire field-initiated research program of the Department of Education supports about 15 to 17 one-year grants.

Foundations, which have the power to spark innovation through funding priorities, tend to fund "action projects" that can be widely franchised with limited research on comparative effectiveness, evaluation, and affordability.

According to a National Academy of Education survey, less than three percent of grants form the major foundations are targeted for educational research. Only about 15 percent of grants categorized specifically as education-related go to educational research.

Recommendations

The National Academy for Education makes a series of policy recommendations for a major effort to strengthen research to better help our education system meet national priorities. Following are some key recommendations:

Funding and Support

—In the next few years, support for research should be increased from one-half a percent of total elementary and secondary school expenditures ($125 to $150 million) to 1 percent of all expenditures for education ($250 to $300 million). Many knowledge-producing industries spend anywhere from 4 to 6 percent of their operating budgets on research and development.

—The NAE supports the notion that much of educational research be funded through mission-oriented institutes. We must select a smaller number of serious education problems and address them in a systematic fashion over a sustained period at reasonable funding levels. The current one- to two-million dollar funding levels for centers are barely enough for staffing and overhead and is not enough to provide support for the necessary long-term experiments, monitoring, and follow-through.

—Research dollars should be centered around the Academy's research priorities that envision the future and are likely to lead to positive change in schools.

QUALITY CONTROL

—We need to develop a National Panel of Reviewers comprised of education, business, and political leaders to advise the federal R&D effort, proposing consensus on what is known and recommending new studies to close gaps in the research base.
—A better balance must be achieved between research from federal centers and regional laboratories and research initiated in the field, and more funds should be available for site visits to monitor field-initiated studies.

INCENTIVES FOR RESEARCHERS

—New incentives are needed to draw talented young people into educational research, including scholars from disadvantaged and minority backgrounds.
—The government must provide increased support for field initiated research, which could be coordinated by Federal research centers in order to take advantage of existing mechanisms for accumulation of knowledge and dissemination.

RESEARCH AND PRACTICE

—To promote greater connections between research and practice, clear incentives are required if researchers are to be engaged seriously with educational practitioners. Incentives are needed for school districts and for state education agencies to collaborate on research and development.
—In addition to centers, labs, and field-initiated research, incentives should be given to states and districts to mount coordinated efforts with scholars and districts. Policymakers should consider ways to integrate research, development, and practice in program planning.

PREPARED STATEMENT OF THE NATIONAL SCHOOL BOARDS ASSOCIATIONS

APRIL 1991 INTERIM REPORT ON THE EVALUATION OF THE 1990 NAEP TRIAL STATE ASSESSMENT

NATIONAL ACADEMY OF EDUCATION

PANEL ON THE EVALUATION OF THE NAEP TRIAL STATE ASSESSMENT PROJECT

APRIL 1, 1991

EXECUTIVE SUMMARY

Three years ago, Congress, the Administration, the nation's Governors, the Chief State School Officers, and other education professionals determined that the time had come to see whether the National Assessment of Educational Progress (NAEP),the "Nation's Report Card", might also become a report card for the states. In the spring of 1988, Congress enacted P.L. 100-297, authorizing a NAEP Trial State Assessment (TSA) program to determine whether state assessments following the NAEP format could produce reliable and useful estimates of educational progress. As part of the authorization, Congress called for an independent evaluation of "the feasibility and validity of [state] assessments and the fairness and accuracy of the data they produce." The evaluation was to be "conducted by a nationally recognized organization (such as the National Academy of Sciences or the National Academy of Education)."

Congress viewed an evaluation of a set of trials in the states as prerequisite to the establishment of a NAEP program at the state level. Major issues to be investigated included the reliability and validity of the data yielded by testing a representative sample of a state's students; the utility of an indicator system, such as NAEP, for guiding state policy; and the effects of state NAEP, positive or negative, on national NAEP. All in all, Congress wanted to estimate the range of benefits of expanding NAEP, in light of its potential cost.

The evaluation of the TSA is being carried out under a grant from the National Center for Education Statistics to the National Academy of Education. To conduct the evaluation, the Academy appointed an independent Panel, co-chaired by Professors Robert Glaser and Robert Linn. Its first mandated report will be delivered to

the Acting Commissioner of the National Center for Education Statistics in October 1991, with the purpose of providing results of the Panel's evaluation of the 1990 trial to Congress, the participating states, and the Executive Branch. The Panel has chosen to issue an interim report at this time for two reasons. First, the authorization for the TSA runs out in 1992, and it is the Panel's understanding that reauthorization hearings may begin soon. Second, the Panel believes that Congress might find its preliminary conclusions and recommendations about the 1990 trial useful, given current attention to the role of assessment in improving educational performance.

Because the first TSA results will not be released until June 1991, it is too soon to evaluate many aspects of the trial, including the various uses and impacts of the results. However, the Panel's preliminary research and deliberations provide the basis for making a set of recommendations to Congress, the states, and the Executive Branch. The Panel believes that these recommendations can help inform the decisions Congress will soon make concerning reauthorization of state NAEP. Justification for the recommendations is presented in the attached full interim report.

Thus far, the results suggest that the 1990 trial has gone well. The Panel has not discovered any significant flaws in the sampling or administration procedures that would threaten the integrity of the results. Nor has it discovered indications that the TSA has adversely affected the national assessment. Consequently, in its role as independent evaluator of this important initiative, the Panel offers the following recommendations:

1. On the basis of its preliminary findings from the 1990 trial, the Panel recommends the release of the state-level 1990 NAEP mathematics scale scores as scheduled.

2. The Panel recommends that future authorizations for state NAEP include adequate resources to sample private school students in order to increase the comparability of results from one state to another, as well as comparability to the national assessment sample.

3. Because of serious concerns about the validity of the achievement levels developed last fall by NAGB, the Panel recommends that NCES arrange for an independent technical review of NAGB's ongoing replication and validation studies, prior to adoption, use or reporting of achievement levels.

4. The use of NAEP at the school district or school level should be authorized only after careful review of policy, technical, logistical, and cost factors. The Panel plans to review such factors and recommends that the prohibition on the use of NAEP scores at the school district or school levels remain until such a review is completed.

5. Because only two subjects at grade 4 and one subject at grade 8 will have been assessed at the conclusion of the 1992 TSA, the Panel recommends the continuance of the trial program In 1994, rather than the full establishment of a state NAEP program. Specifically for 1994, the Panel recommends trials at three grade levels—fourth, eighth, and twelfth—in mathematics, reading, and one additional subject, such as science.

6. Substantial lead-time is required for achieving national consensus on new content frameworks, and for developing assessment questions and exercises that elicit more than rote learning from students. Therefore, the Panel recommends that authority for continuation of state NAEP be made at the earliest time and that Congressional appropriations be at a level that will support appropriate assessment innovations.

These recommendations are offered in hopes of contributing to a thorough evaluation of the promise of state NAEP. As state-level trend lines are established for achievement at various levels, in various subjects, the Panel anticipates that TSA data can become increasingly valuable to the participating states. However, the Panel wishes to register here, in addition to these recommendations, a caution against the overinterpretation from TSA results to judgments about causes or explanations of group differences in achievement. In particular, it would not be warranted from the NAEP data only to conclude that higher scores are the result of any particular differences' in state policies or educational practices. As the trials move forward, it will be essential to the long-term effectiveness of this venture that those who use NAEP data exercise caution and avoid unwarranted interpretations.

Ambach dissents from this position; he is on record elsewhere as recommending lifting the prohibition at the school district level where the size of the enrollment enables sampling as used at the state level.

April 1991 Interim Report on the Evaluation of the 1990 NAEP Trial State Assessment

National Academy of Education Panel on the Evaluation of the NAEP Trial State Assessment Project

April 1, 1991

For more than twenty years, the National Assessment of Educational Progress (NAEP) has been the best available indicator of the status of the nation's educational system. Unlike results from college admissions tests that are often used inappropriately as indicators of nationwide educational achievement, NAEP represents all students, not just a subset of college-bound high school seniors. NAEP trend data have shown that, from 1969 to the present, the average overall achievement levels in the core disciplines of reading and mathematics have been quite stable; however, achievement levels for too many students are below the levels required for their successful participation in the workforce and for the well-being of the nation. Of particular concern is the performance of 17-year-olds in science, where there has been a significant decrease in achievement over the past twenty years. Only a small proportion of students attain the basic scientific knowledge needed in this society; most fall behind early in learning science.

NAEP has been the source of some encouraging information as well. It has provided valuable insights into variations in achievement by race, ethnicity and gender. Through NAEP, policymakers learned in the 1970's and 1980's that minority students had begun to narrow the gap between their academic achievement and that of whites–though that gap remains unacceptably large.

NAEP's role as an independent indicator of educational progress is quite different from that of tests that supply information for school accountability or measure an individual student's achievement. NAEP's role is unique in that, since its first administration in 1969, it has provided the most reliable single source of information about trends in the achievement of the nation's youth. Although we may not like the discovery that levels of achievement have changed relatively little during the past twenty years and remain below those to which we aspire, NAEP will allow us to continue to monitor progress for the nation as a whole as we renew efforts for improvement and reform.

Assessment, of course, has other roles as well. The current national debate about establishing a national examination system or a national test of individual students' performances centers on using tests that would be integral to state curricula and address standards of achievement. It is critical, however, that the purposes of a national or state level indicator system be clearly distinguished from those of individual tests. NAEP was not designed to provide scores for individual students or schools. Indeed, such uses are precluded in the current law, which bars student identification and the reporting of results for individual schools. Ranking, comparing, or evaluating individual students, schools, or school districts is also prohibited. In the context of current ambitions for educational change, such as a proposed national examination system and school restructuring, NAEP is best seen as an indicator that can reflect the outcomes of these changes.

Trial State NAEP

In 1986, Secretary of Education William Bennett formed a study group to look at NAEP and to suggest ways to improve the process for assessing student achievement in the United States. Our new Secretary of Education, Lamar Alexander, served as chairman. The study group's document, *The Nation's Report Card* (prepared by Alexander and H. Thomas James, President Emeritus of the Spencer Foundation), recommended expanding NAEP to provide baseline and trend achievement data for the states. This recommendation was consistent with growing interest in educational progress at the state level. The report noted that primary responsibility for education in the U.S. historically has been vested in the states and argued that the value of NAEP would be enhanced if it reported state results. Participation in a voluntary state NAEP program, the report further argued, would preserve local educational autonomy and, at the same time, give states access to a core of high-quality data on performance.

The enactment of Public Law 100-297 in the spring of 1988 provided for the voluntary participation of states in NAEP on a trial basis in 1990 and 1992. In February of 1990 the first trial of the state NAEP, an assessment of mathematics achievement, was administered in more than 3,500 schools, to some 100,000 of the nation's eighth graders. In total, 37 states, the District of Columbia, and two territories par-

ticipated, an indication of the wide interest in state NAEP. The results are to be released on June 6, 1991.

The second trial is scheduled for 1992, with expanded data collection to include fourth grade reading as well as fourth and eighth grade mathematics. The continuation of the trials through 1994, however, is contingent on Congressional action.

The Panel's First Year of Activity and the Reasons for this Report

Public Law 100-297 also mandated that an independent evaluation be conducted to assess the feasibility and validity of the Trial State Assessments (TSA). In October 1989, the National Center for Education Statistics (NCES) commissioned the National Academy of Education (NAE) to conduct this evaluation. The NAE assembled a panel of experts in a broad range of technical and policy fields in education and arranged for technical and staff support from the American Institutes for Research (AIR). The Academy's panel held three meetings in its first year. During this time, the Panel focused on the data and information it would require, and the design of studies needed to conduct an effective evaluation. The Panel has made major decisions on a first-phase agenda for obtaining information about the 1990 trial, its impact, and questions related to the future value and validity of the TSA.

The key questions that guide the Panel's work are those Congress, in 1988, anticipated would be crucial to evaluating the TSA: (1) How well was the assessment implemented from a technical perspective? (2) How valid and accurate is the assessment? Has it yielded valid and reliable data at the state level? (3) How useful are the results and reports generated from the assessment? To answer these questions, the Panel has commissioned a set of studies and papers to address prominent aspects of the TSA design and implementation. The results of these inquiries will clarify the appropriate role of state indicators. They will also reveal how state NAEP might help monitor progress toward national educational goals.

Because authorization hearings beyond 1992 may begin soon, the Panel has chosen to issue this interim report, prior to releasing its first mandated report in October. Although it is still early in the evaluation, the Panel's findings and discussions to-date have direct bearing on issues that may be considered in connection with future authorizations. These findings can also inform national deliberations about the use and effects of educational assessments, particularly those relating to the work of two groups: the National Education Goals Panel and the President's Educational Policy Advisory Committee.

The balance of this interim report includes the following sections: A review of the Panel's work, along with its evaluation of the trials to this point in time; discussion of the achievement levels established by NAGB; discussion of the prohibition against reporting NAEP results below the state level; suggestions regarding the reauthorization of state NAEP; and a short overview of topics for which data will be available for the October report.

The Evaluation of the Trial To-Date

At present, the Panel is prepared to offer preliminary observations about the 1990 TSA based on data and deliberations in four areas: (1) sampling, (2) excluded student populations, (3) administration, and (4) inferences that can and cannot be made from the 1990 Trial State Assessment. A more complete and detailed report on work in these areas will be presented for Congressional, state, and Executive Branch consideration in October.

Sampling

The preliminary analysis of the sampling design and its execution has focused on the reports on sampling in the February 1990 trial that were available as of January 1991. From this analysis the Panel has concluded that the sampling was competently performed. A common difficulty encountered in programs such as NAEP is nonparticipation; some schools refuse to participate and some students either refuse or are absent. The magnitude of nonparticipation in the TSA as a whole was reasonably small, with about 6 percent of the schools declining to join the project and about 6 percent of the students in the participating schools not taking part. These rates varied from state to state, however, and in two states the rate of students not participating was between 10 percent and 20 percent. Statistical adjustments, known as "nonresponse adjustments," are being used to compensate for the missing data; the adjustments appear to be appropriate and reasonable. In sum, school and student participation in the 1990 sample produced a generally favorable picture for state NAEP.

BEST COPY AVAILABLE

In considering issues of sample design, the Panel has been alert to any indication that state NAEP might have a negative impact on participation in national NAEP. Thus far we have found no cause for concern. While it is true that state NAEP did increase the burden on small states to provide a sufficient number of schools to meet the requirements of both the national and state sampling frameworks, few of the 37 states that participated in the first trial found it to be a problem. Furthermore, there was no indication in the administration of the 1990 trial of interference with the administration of national NAEP. In sum, the Panel can report that, thus far, the 1990 TSA has had no discernible negative impact on the 1990 national NAEP.

Excluded Student Populations

The design of the TSA allowed for the exclusions of three groups of students: students enrolled in private schools, students with limited English proficiency (LEP), and special education students with individualized education plans (IEP). Together, these groups make up about one-sixth of the eighth grade nationally. Differences among the states in the proportions of students in these groups could have important effects on state-by-state and state-to-national NAEP comparisons. For example, in national NAEP, which tests both private and public school students, private school students tend to perform better than public school students. Because comparisons of states' performances on NAEP inevitably will be made, the exclusion of private school students in the TSA is cause for concern; states' performances could change substantially with inclusion of the private school students.

Private School Students. Private school students typically made up the largest excluded group in each state. Nationally, about 12 percent of the eighth grade students are enrolled in private schools, but this percentage varies widely across states: In seven states fewer than 5 percent of the eighth grade students are in private schools, and in seven others more than 18 percent are in private schools. Wyoming and Utah each enroll only about 2 percent in private schools, whereas Hawaii and the District of Columbia each enroll about 20 percent.

At this time the Panel does not know how much the inclusion of private school students would affect the rankings of states. The magnitude of the effect depends on how many students in a state are enrolled in private schools and on the size of the differences between public and private school students' performances on the NAEP items. For the October report, the Panel is conducting analyses to examine how the states' results might change as a function of the exclusion of private school students.

The Panel believes that state NAEP data would better reflect educational achievement and make state results more readily comparable if, in addition to results for public school students, results for all students (in public and private schools) were produced.

Limited English Proficiency (LEP) and Individualized Education Plan (IEP) Exclusions. The exclusion criteria for LEP and IEP students were part of the sampling design and were implemented locally, but the local administrators were told to include doubtful cases in the assessment. Overall, about 1 percent of the students were excluded for reasons of limited English proficiency and about 4 percent because they had individualized education plans. But the percentages across states varied, with LEP exclusions ranging from near 0 percent in West Virginia and about 2 percent in New Jersey, New York, Rhode Island, and Texas, to 5 percent in California. IEP exclusions ranged from about 2 percent in Montana to 8 percent in Arkansas.

The Panel is in the process of examining how consistently the exclusion rule was implemented in the first trial state assessment, but findings will not be available until the Panel's October report.

Administration

As part of the study on administration, Panel staff directly observed training sessions for test administrators and assessment sessions. In addition, they conducted independent analyses of the TSA data and Quality Control Monitoring Data collected by the NAEP contractor. Finally, they conducted a survey of State Testing Directors as an independent appraisal of the 1990 TSA administration. Their initial findings focus on issues of local conditions for implementing the TSA, the impact of the first trial on the 1990 national NAEP, and state testing directors' responses to the first trial.

Local Conditions. Unlike national NAEP, the TSA employed local school staff to administer the test sessions, rather than staff employed by the contractor. Uniform-

ity of assessment conditions is a prerequisite for the legitimate comparison of a state's results, both with the national composite result and with the results of other states. The administration of TSA by local staff had the potential to threaten the comparability of state results with national results. Although the local test administrators underwent careful training, there was the possibility that students might respond differently when tested by them, and that this would distort the results of the TSA. Therefore, the NAEP contractor had a monitor present in a random half of the test sessions to assure that the local administrators proceeded according to their training.

The critical finding was that student performance in monitored sessions did not significantly differ from the performance in unmonitored sessions, suggesting that local administrators were largely successful in implementing uniform testing conditions that did not advantage or disadvantage students. Quality control monitors looked for variations in every aspect of the testing session, including timing, reading the script, and handling student questions. Analysis of the reports indicates that deviations from uniform procedures were infrequent and were unlikely to have systematically influenced state results.

The Trial and National NAEP. Because of possible differences in testing conditions, a second issue was whether students assessed in the TSA might obtain scores that, on average, differ from those obtained by students in national NAEP. Since the sampling frame for the TSA differed from that of national NAEP, the contractor constructed a "matched" subset of the students from national NAEP to enable valid comparisons. Compared to the matched subset of national NAEP students, students in the TSA obtained slightly, but reliably, higher scores. While the design of the study prohibits a definitive explanation for the difference, the Panel is exploring the possibility that students participating in state NAEP may have been more motivated to do well than those in national NAEP. The Panel will report further on this issue in its October report. In addition, the Panel will closely monitor the 1992 trials to see if this potentially important finding is replicated.

TSA Planning and Policy from the State Testing Directors' Perspective. State testing directors are knowledgeable observers and important stakeholders in the assessment prccess. To monitor their responses to the first trial, the Panel staff conducted an independent survey of the directors as part of the study on administration. Most reported that the assessment went well and that the data from the TSA would be of value to their states. However, some noted that they felt excluded from important policy decisions in the establishment and implementation of the TSA. The Panel applauds the efforts of NAGB and NCES in their stated intention to use CCSSO's Education Information Advisory Committee as a vehicle for providing state testing directors with greater policy input. However, because the Panel recognizes the need for close cooperation among NCES, the contractor, and the states participating in the TSA, the Panel proposes that the governance and administrative structures of NAEP strengthen the mechanisms for securing input from state testing directors into the state NAEP policy and assessment development process.

Based on the preliminary results of its studies and its ongoing deliberations, the Panel believes that the 1990 TSA has proceeded well. Thus far, the studies have identified no signs that the experiment is flawed, that major redirection is necessary, or that the TSA should be terminated. On the basis of its preliminary findings from the 1990 trial, the Panel recommends the release of the 1990 NAEP mathematics scale scores as scheduled.

Despite this generally favorable observation, the Panel is concerned that the exclusion of private school students from the TSA ultimately will diminish the utility of the trials and future administrations of the state NAEP. Given both the significant variation from state-to-state in the size of this group, and its inclusion in the national sample, issues of comparability become much more complex than need be when private school students are excluded from the sample. The Panel recommends that future authorizations for state NAEP include adequate resources to sample private school students in order to increase the comparability of results from one state to another, as well as comparability to the national assessment sample.

Inferences That Can and Cannot Be Drawn from the Trials

Congress should be aware of the kinds of inferences that can be usefully drawn from the TSA, given the design of the assessments. As state-level trend lines are established for achievement at various levels, in various subjects, state NAEP data will become increasingly valuable to the participating states. They will provide governors, legislators, and state school officials with the ability to monitor educational progress using information of unparalleled richness. These trend lines will enable

comparisons with similar states, the nation, and other countries as the basis for much-needed educational innovation. The two data points for eighth grade mathematics provided by the 1990 and 1992 TSA's will provide valuable preliminary trend information to those states that participate in both trials. However, the real value will come with the accumulation of additional data points across time.

The ability to compare similar states will prove useful in the consideration of policy issues. It should be emphasized, however, that the results will not support causal inferences about what produced differences in achievement. In particular, it would be safe to conclude that higher scores are the result of any particular differences in state policies or educational practices.

At this juncture, it is important to remember that the 1990 Trial State Assessment is limited in scope: it embraces only one subject at one grade level, eighth grade mathematics. With the inclusion of fourth grade reading, fourth grade mathematics, and eighth grade mathematics in the 1992 trial, policymakers and the public will have a valuable, yet narrow, window on learning outcomes across the two grade levels and curriculum areas. The Panel cautions against overgeneralization from these trials to questions of schools' and teachers' performances or group differences in achievement.

The results will see their best use in the establishment of trends in achievement within a state, over time, and in the drawing of comparisons between states with similar populations, and between a state and the nation. However, comparisons of states' rankings inevitably will be made. While states can be ranked with respect to mean levels of achievement, interpretations of state-to-state differences must be made with great caution. Three issues must be addressed. (1) It must be determined whether the differences between the rankings are large enough to be considered reliable. (2) The relevance of a state's ranking to judgments about its educational quality will depend upon the match between the content tested by NAEP and the state's curriculum framework as implemented. Some state frameworks are closer than others to the content of NAEP. (3) Differences in states' performances may be due to differences in demographics. The Panel has studies in place to examine all three of these issues.

Issues Currently Under Discussion and Debate

Since the Panel received its mandate for the evaluation of the 1990 TSA from NCES, two important policy issues relevant to state NAEP have become prominent. The first is the proposal for and the development of a set of achievement levels or standards, using the 1990 NAEP mathematics items. The second is a recommendation by the National Assessment Governing Board for lifting the current prohibition against the reporting of NAEP results below the state level. Given the importance of both these issues for state NAEP, the Panel has agreed to address them in this report.

"Standards" or Achievement Levels

The legislation that authorized State NAEP (P.L. 100–297) also assigned to NAGB the task of developing appropriate achievement goals for each age, grade, and subject area in NAEP. The unveiling of a set of six educational goals by the White House and the Governors in 1989 heightened interest in educational standards, and set the stage for NAGB to develop a set of achievement levels that could be used to measure progress toward the national goals. Last August, NAGB engaged in an exercise to define basic, proficient, and advanced achievement levels in fourth, eighth, and twelfth grades, using the 1990 NAEP mathematics items. The Panel applauds this attempt to make scores more interpretable, but cautions that it must be viewed as an intricate process involving judgment, definition, and, ultimately, issues of reliability and validity.

As valuable as achievement levels might be for the states in monitoring their progress toward meeting some of the national educational goals, the results of the process of setting the achievement levels should meet the scrutiny of experts and be credible to the public. The Panel concurs with NCES's Technical Review Panel and CCSSO that the current achievement levels, obtained before January 1991, are flawed. As a result, the Panel's Chairman and Co-chairman have written to Richard Boyd, Chair of NAGB, urging that the achievement levels be used only if corrected. NAGB is in the process of conducting a replication and validation study in four regions of the country. The Panel commends NAGB's efforts to secure validation of the achievement levels, since the data collected for that purpose should be adequate for evaluating the current levels, or if necessary, modifying or discarding them. Since the Panel believes that the use of inadequately developed achievement levels

could have a corrosive effect on state participation in the future, as well as on the credibility of NAEP more generally, the Panel will monitor the validation studies.

Reporting TSA Data Below the State Level

NAGB recently has recommended to Congress that the current prohibition against reporting NAEP data below the state level be lifted to allow reporting at the school district or school. NAGB would continue the prohibition on reporting individual student scores. The Panel supports NAGB's recommendation to continue the prohibition against reporting data at the student level. But the Panel also believes that expansion of NAEP to provide results at the individual school building level or for other than large school districts could lead to the loss of NAEP as an independent and uncorrupted indicator of educational progress. NAEP's historic role as an auditor that stands apart from the training and testing of individual students can too easily be compromised by its use at the school and student level.

The extension of NAEP to the district level raises a somewhat different set of issues. The reporting of data for at least some of the largest districts may be as warranted as reporting data for some of the smallest states. Indeed, because of its special status, the District of Columbia did participate in the 1990 TSA. Prior to lifting the prohibition, however, the Panel believes that the technical, policy, and cost implications, as well as the implications for future test design and administration, need careful study and consideration. The Panel plans to commission a study on the implication of reporting NAEP data below the state level and will present the results and conclusions in a future report. The use of NAEP at the school district or school level should be authorized only after careful review of policy, technical, logistical, and cost factors. The Panel plans to review such factors and recommends that the prohibition on the use of NAEP scores at the school district or school levels remain until such a review is completed.

Panel Perspectives on Key Issues in Reauthorization

The Panel recognizes the great value of maintaining continuity of state NAEP, especially in light of the general technical success of the 1990 trial. But as Congress considers reauthorization of 1994 NAEP, the Panel suggests a number of important issues to consider.

Planning the 1994 Trial State Assessment

The Panel recommends that 1994 NAEP, when reauthorized, should include additional state trials since, with the conclusion of the 1992 trial, only two subjects, mathematics and reading, will have been evaluated at two grade levels. In 1994, national NAEP will assess mathematics, reading, science, and history and geography combined. Authorizing state trials for one subject (e.g., science) in addition to reading and mathematics and for an additional grade level (twelve in addition to four and eight) prior to moving to a fully implemented state NAEP would be informative. By 1994, trends for fourth grade mathematics and reading would be available in addition to the trends for eighth grade mathematics for 1990 and 1992, thereby allowing for a more complete evaluation of the uses of and the interest in such trend data by the participating states. In addition, such an expansion would provide data to help evaluate the feasibility, impact, and cost of a fully implemented state NAEP.

The Panel suggests the addition of a twelfth grade trial in 1994. Of central importance to the Panel is the fact that results from the trials at the fourth and eighth grade levels cannot be assumed to generalize to the twelfth grade. The motivation of twelfth graders to participate and perform well may be very different from that of students in the lower grades. Moreover, state level results for twelfth grade students may be of particular interest and use to the states. There is great concern about workforce preparedness on the part of private industry, the Administration, Congress, and the states.

Finally, preliminary evaluation results suggest that the 1990 trial is going well. However, before the Panel can reach a final conclusion regarding the success of the trials, it must complete its evaluation of the 1990 and the 1992 trials. There is much useful information to be gained from continuing the trial program to inform the fuller development and implementation of state NAEP in the longer term.

Because only two subjects at grade 4 and one subject at grade 8 will have been assessed at the conclusion of the 1992 TSA, the Panel recommends the continuance of the trial program in 1994, rather than the full establishment of a state NAEP program. Specifically for 1994, the Panel recommends trials at three grade levels—

fourth, eighth, and twelfth—in mathematics, reading, and one additional subject, such as science.

Assuring the Quality of State NAEP

With Congress' requirement in 1988 that a national consensus process be carried out when updating test content frameworks, NAEP has reaffirmed its status as an innovator. The 1992 reading assessment reflects the current emphasis on performance-based assessment, and the 1994 science assessment seems likely to pursue the same progressive route. The Panel believes that NAEP should exemplify and promote current innovations in assessment technology on a stage-by-stage basis. To provide for trend data, provision must be made for assessments to include items that maintain links to past assessments and, at the same time, build links to the future. The consensual development and updating of content frameworks are essential to securing innovation and planning this balance.

New assessment technologies and innovations carry with them increased costs and require considerable time to develop. The Panel is also aware of the massive amount of work that must be completed in relatively short periods in the implementation of NAEP. For example, CCSSO had less than four months in 1989 to create the reading framework and must, in seven months, create the new science framework. Working within such schedules, while incorporating high-quality innovations in assessment technology, is nearly impossible.

Substantial lead-time is required for achieving national consensus on new content frameworks, and for developing assessment questions and exercises that elicit more than rote learning from students. Therefore, the Panel recommends that authority for continuation of state NAEP be made at the earliest time and that Congressional appropriations be at a level that will support appropriate assessment innovations.

The Panel's October Report

The Panel's mandated report in October will expand on the topics addressed in this interim report, and will focus as well on: the presentation and impact of the results of the 1990 Trial, the content validity of the items, and the policy context of goals for achievement in which the TSA is embedded.

The results of the 1990 Trial State NAEP will be released on June 6, 1991. The Panel is interested in the clarity, interpretability, and usefulness of different formats for reporting results to the states. It will also investigate any moves toward curricular or instructional changes in states' mathematics programs. Finally, it will examine the degree to which the reports are fair–that is, the degree to which the rankings of states vary as a function of different types of test content (e.g., algebra versus geometry), or as a consequence of adopting alternative methods for producing an overall score. The Panel will also examine the relation between state assessment results and the racial, ethnic, and gender composition of the states.

When Congress authorized NAEP in P.L. 98-511, it required that the curriculum frameworks be developed through a national consensus process, providing for the participation of teachers, curriculum specialists, school administrators, parents, and members of the general public. In October, the Panel will report on the adequacy of this consensus process for the 1990 mathematics assessment and the 1992 reading assessment. The report will describe the constituencies represented and the nature of the advice sought. It will also evaluate how this advice and input affected the design of the frameworks, and the extent to which the frameworks represent a consensus among professionals in the fields of mathematics and reading education. Of particular interest for TSA is the degree to which the consensual process represents a national perspective that includes the current goals and objectives of state and local school districts.

There is considerable interest in using the results from the achievement levels for inclusion in the state "report cards" that the National Goals Panel will release this September. In addition, discussion continues about whether there should be a national examination, and if so, what role NAEP and state NAEP should play if a national examination is established. The Panel continues to monitor the policy context in which the 1990 Trial is occurring, and will report more fully on that context in its October report.

The Panel hopes that this interim report regarding the 1990 trial in mathematics and TSA reauthorization will prove useful to Congress as it deliberates about the future of state NAEP. The recommendations endorsed here will allow thorough evaluation of its promise as a valuable indicator of states' educational achievement and will strengthen the possible full extension of NAEP to the states. Over the shorter and longer terms, state NAEP may serve as a vital measure of progress

toward the achievement of the educational goals that are a priority for the states and the nation.

Ambach dissents from this position; he is on record elsewhere as recommending lifting the prohibition at the school district level where the size of the enrollment enables sampling as used at the state level.

Prepared Statement of The National School Boards Association

I. INTRODUCTION

The National School Boards Association (NSBA) represents the 97,000 local school board members responsible for educational governance in the nation's 15,350 public school districts. We are pleased to provide this statement for the record regarding the reauthorization of the Department of Education's Office of Educational Research and Improvement.

NSBA believes the issues of national standards and testing involve some of the most important decisions facing public education today. These issues have special relevance for local school districts because they are closely tied to the schools' accountability for learning and what is actually being taught in the classroom. Testing, in particular, is of high concern to schools because of the profound impact it can have on education performance and the future of millions of school children.

Our statement focuses particular attention on issues relating to the reauthorization of the National Assessment of Education Progress (NAEP). NSBA believes that lifting the prohibition against local use of NAEP at this time would be premature and potentially damaging to school improvement efforts. Any movement toward using NAEP to report school district test results must be made deliberately based on solid data from evaluations of state, and then local trial results. There is no compelling reason to alter language in the 1988 Hawkins-Stafford Act prohibiting local use of NAEP until this data is available and fully appraised. Many other leading voices in education—including groups representing teachers, administrators, principals, parents, and the civil rights community, as well as assessment experts—share this judgment. On the other hand, if the issues raised in our statement below are carefully weighed and adequately addressed, it is possible for NSBA to support a system of national standards and testing used locally in which NAEP could have a significant role.

As additional background, we have included a copy of a letter sent to Governor Roy Romer, Chairman of the National Education Goals Panel, further outlining NSBA's position with regard to national, state, and local testing issues.

II. NSBA'S POSITION ON REPORTING NATIONAL, STATE, AND LOCAL TESTING

A. National Reporting: Valued Activity

The National School Board Association (NSBA) has historically supported the vital function of the National Assessment of Educational Progress (NAEP) as our only nationally representative assessment of what America's students know and can do. We believe the value of NAEP over its twenty year life has been its ability to develop a national baseline of information about performance and trends in learning in a manner that has provided policymakers with the ability to make substantive improvements in education.

B. State Reporting: Promising Function

In 1988, NSBA supported Congress' decision to permit four trial state assessments (1990 eighth grade math assessment; 1992 fourth and eighth grade math assessments, and fourth grade reading assessment) to gauge whether information deemed so valuable nationally could be reported at the state level as well. Our support was based on specific language in the statute authorizing the National Assessment Governing Board (NAGB) to undertake trial state assessments for the purpose of providing states with the kind of information needed to identify problems and develop solutions for improving educational performance. Included in the statute was explicit language specifying that each trial must be carefully monitored and independently evaluated and appraised. NSBA is looking forward to an evaluation of the recently released math assessment, as well as the three trials that will follow in 1992. Only

after these evaluations are available in 1993 would Congress have adequate information to determine whether further extension of NAEP testing would be productive.

C. Local Use of NAEP: Premature

NSBA commends NAGB, its contractors, the Department of Education, and others involved in the state trial effort to develop fair and honest state assessments. However, because this review is still in process, and in light of a number of other important issues and concerns raised below, we believe it is premature to extend NAEP, or any other national test, to the local level. Our position is based on several reasons:

1. Before local testing of any kind ensues, it is critical to establish national standards for education so that we understand clearly what is to be tested—as well as the purpose of the test. A ten-year process has been established, involving Congress, the President, and the nation's governors and school policymakers, to do just that under the umbrella of the National Education Goals Panel. Therefore, the recommendation to allow local use of NAEP for reporting test results is out of step with both the preferred approach of setting standards first, and the sequence of events which are unfolding nationally.

2. The price paid for developing a local NAEP which is out of sequence with national education standards unnecessarily invites serious misdirections in policymaking. The potential negative impact can be measured in terms of (a) public confusion and lowered confidence in the schools and policymakers, (b) the financial cost burden for school districts in administering and acting upon misdirected local test results, (c) the influence on curriculum and course offerings, and (d) the influence on innovation, instructional practices, and issues relating to cultural diversity. By contrast, a properly sequenced local report would provide greater assurance that the test will be beneficial and in context with the ten-year plan for educational improvement now being developed.

3. Even if standards are put in place and NAEP is the instrument chosen to measure them, there is still not enough experience with state data (although the NAEP 1990 trial state math assessment is promising) to know how it would be used locally. Moreover, there has been no national discussion as to how local reporting should be accomplished and what information the local report should contain.

4. No evidence presently exists answering concerns of local school policymakers that local NAEP reporting will produce more than a comparative ranking of school districts by states, and cast unfair blame upon schools for "not measuring up," particularly as new world standards are adopted. in other words, the same issues that caused Congress to prohibit local reporting in 1988 may still persist today.

5. Last month, NAEP's report on state math programs could not avoid invidious comparisons based on state rankings. Unfortunately, the data from the trial has been reported in a manner that has caused the public to draw conclusions based on such rankings. Until the public understanding of NAEP is changed, similar reporting of test results at the local level may force political considerations to overtake sound educational policymaking.

6. NAGB's recommendation that states and school districts be permitted to use NAEP at their option only begs the questions which we have raised; it does not resolve them. Local school districts would he left unprotected against federal or state desire to mandate, coerce, or otherwise require local reporting. Moreover, the premature commencement of local reporting of any one test, even optionally, could create a presumption that would unnecessarily bias the standard assessment system which is developed.

III. CRITERIA FOR TESTING AT THE LOCAL LEVEL

Despite these issues, a growing consensus has emerged recognizing that results of well-conceived national testing can be beneficial for school districts if they include constructive information and accountability to foster educational improvement. In this regard, it is very encouraging that the 1990 NAEP trial assessment includes score ranges and interpretive and underlying data to identify trends in achievement.

NSBA believes that a national test, including one that provides local reporting, should adhere to the following guidelines: (a) the test should be tied to students' mastery of skills but not specific subject matter (i.e., assuring flexibility in local curriculum decisions), (b) student performance should be reported in terms of ranges of mastery, not specific point scores; and (c) performance should be reported in a manner that avoids generalized conclusions implying inherent differences in ability on the basis of race, national origin, or gender.

Additionally, before lifting the prohibition against local NAEP reporting, we believe the following issues must be addressed:

A. Any local use of NAEP must have as its primary purpose the goal of assisting school districts to improve education and advance student achievement. The value of such testing lies in providing useful information to school districts about their educational strengths and weaknesses. Aggregated with other noneducational indicators and context factors, information of this kind can give school policymakers and practitioners a more comprehensive assessment through which to identify causes, potential solutions, and resources in the name of improved educational performance.

B. A local NAEP reporting process that does not aptly reflect the diversity of schools and becomes a method to cast winners and losers through pointless rankings of schools and school districts will be of limited value. Any local assessment must be sensitive to cultural, social, class and gender differences, and to disabilities. It should be valid for all groups and not penalize any groups.

C. The use of local NAEP reporting, particularly in conjunction with nationally developed standards to meet the education goals, should not unnecessarily dictate curriculum, stifle instructional innovation, or fail to adequately recognize cultural diversity at the local level.

D. No matter how reliable the results, no matter how enriching the assessment data, if local reporting occurs in isolation from other resource and contextual issues, the entire reporting process will be subject to broad-based challenge by local school policymakers. We believe that a thorough appraisal of interpretive input" factors is essential to ensure an honest assessment. That appraisal should identify educational resources available to the school district, such as expenditures per student, availability of materials, equipment, and other instructional technologies. It should identify context factors, such as school climate, type of community, and socioeconomic status and health conditions of students. And it should identify education program information such as school district curriculum objectives, instructional methods, class size, course sequencing, and management strategies.

IV. EXPERT PANELS RECOMMEND CAUTION ON LOCAL NAEP TESTING

NSBA's position on local reporting is supported by two important advisory panels which are providing critical insight to NAGB, as well as members of Congress, the administration, and governors involved in the National Education Goals Panel process to establish national standards:

A. The National Academy of Education's Panel on the Evaluation of the NAEP Trial State Assessment Project

In April, the panel charged by Congress with providing an independent evaluation of NAEP's trial state assessments issued its interim report. It recommended that "The use of NAEP at the school district or school level should be authorized *only after careful review of policy, technical, logistical, and cost factors* (emphasis added). The Panel plans to review such factors and recommends that the prohibition on the use of NAEP scores at the school district or school levels remain until such a review is completed."

The Panel expressed a number of concerns leading to this recommendation. It recognized that although the math trial assessment has gone smoothly thus far, closer examination is necessary of the impact of (a) student populations excluded from the first state trial, (b) whether or not students participating in the state NAEP were more motivated to do well than those in the national NAEP, and (c) what to do to protect against inevitable comparisons of states' rankings. The Panel also cautioned against making generalizations based on an assessment limited to only one subject at one grade level. We believe that these concerns, valid for state reporting, are equally valid for local reporting. The Panel will respond to these issues when it releases its final report to Congress in the Fall.

The report also warned that expanding NAEP below the state level could result in its loss as an "independent and uncorrupted indicator of educational progress. NAEP's historic role as an auditor that stands apart from the training and testing of individual students can too easily be compromised by its use at the school and student level."

B. Interim Council on Standards and Testing

Recognizing the value of involving national, state, and local education policymakers and practitioners, testing experts, and other education stakeholders in the process of developing national standards, Congress and the administration have formed this Interim Council to determine the desirability and feasibility of developing national standards.

At its initial meeting this week, expert witnesses offered provocative insight on standard-setting and testing. Each were clear in enunciating concern over test sequencing, the need to incorporate cultural diversity, the profound impact on local curriculum decisions, the factors implicit in developing high performance, yet reasonable, standards, and other similar issues. They concluded that these must first be resolved before the nation can move forward in developing appropriate measures.

They also discussed the integral, but still unclear, role NAEP will play in this process. For example, it is not known if NAGB's current efforts to use NAEP to identify what students should know—in addition to what they now know—will conflict with the National Education Goals Panel efforts to identify the same. They also questioned how new standards and their measurements (including NAEP) will be used to assure local accountability. And they questioned how long it would take schools to adapt to new standards and measures and what types of adjustments will be needed to compensate for millions of school children now accountable to a different set of "educational standards."

The ultimate conclusion of these experts, and the advice given to the Interim Council, was to move orderly and deliberately in developing standards and measures. Unless the kinds of issues we have raised in this statement are first resolved, the standards will have little value for the nation. We believe these conclusions can be extended to the question of lifting the prohibition against local NAEP reporting.

V. NAGB'S POSITION ON LOCAL USE OF NAEP

During NAGB's testimony before this Subcommittee, reference was made that "some education organizations," are "opposed" to reporting NAEP results at the school district and school building levels. NAGB's testimony stated that the "primary concern of these organizations is that this use of NAEP will permit unwarranted intrusion into local decision making and will tend to establish a national curriculum."

With respect to NSBA, NAGB's written testimony does not adequately reflect the real basis for our concern. As we have stated before—including in testimony before this Subcommittee last year—NSBA is not necessarily opposed to a system of national tests and standards used at the local level for school improvement. While a move toward establishing a national curriculum would naturally be of concern to us, from NSBA's point of view, the unresolved issues we have raised in this statement relating to the use of local testing to improve education are the driving elements that will determine our support.

VI. CONCLUSION

NSBA commends NAGB for the work it is doing on national and state testing. At some point, the use of NAEP tests at the local level may be advantageous. However, given the major impact that such testing would have, we strongly urge that standard setting occur fat. We also believe the prohibition on local testing should be continued for the reasons outlined in our statement, including: (a) there is no compelling reason to alter current statutory timelines authorizing trial assessments and their evaluations; (b) there is not enough experience and data available to warrant local use of reporting of NAEP tests; (c) proceeding with local reporting is out of sequence with the national standard setting process which Congress has just legislated; and (d) there are now few assurances that valuable interpretive input data will be prominently displayed as part of local results that provide districts with credible information to improve educational performance. It may be possible to resolve these issues to allow effective local testing—but we are far from that resolution now. More time and study is necessary to do so.

NATIONAL SCHOOL BOARDS ASSOCIATION
June 10, 1991

Honorable ROY ROMER
Governor of Colorado
State Capitol
136 State Capitol Street
Denver, CO.

DEAR GOVERNOR ROMER: On behalf of the nation's 97,000 local school boards members, the National School Boards Association (NSBA) appreciates your continuing efforts and leadership to advance education through the National Education Goals Panel.

During the June 3rd meeting of the Panel in Washington, D.C., two major points of concern to local school boards arose. First, NSBA agrees with your premise that progress reports should provide the nation with indicators that reflect: (1) where education should be (standards), (2) the current condition of education (assessment), and (3) what needs to be done to get better (strategies for improvement). However, during the discussion several other Panel members expressed the view that progress reports should only reach the second step. That is, they envision a progress report on student performance but do not want to address "input" or strategy-oriented indicators.

During the hearing which you held at the NSBA Convention last April, local school board witnesses repeatedly voiced the concern that unless the availability of resources and the participation of other key sectors (e.g., parents, state and federal governments, and the business sector) were assessed, the goals will not be realized. It is from that standpoint that we urge you to incorporate "input" or strategic indicators into the progress report, including the school readiness indicators which you advocated, as well as the economic indicators suggested by Governor Bayh regarding student testing (Goal No. 3).

Our second concern relates to national testing. We appreciate the remarks you made to President Bush that, while NSBA would not support a single national test, it is open to a more diffuse assessment system. As local leaders elected (or appointed) to govern the schools, we believe that if constructive information and accountability to foster educational improvement can be provided through well-conceived testing, we will have a powerful tool in meeting our policymaking responsibilities.

At the same time, we have several caveats on the testing issue. For example, the national standards selected and the decentralized assessment system utilized should, among other criteria, be designed to indicate mastery of learning skills and subject matter, not to establish curriculum. Equally important, test results (including NAEP) should be aggregated, and presented in a manner that produces constructive change—rather than rankings by which to censure those at the bottom end. We believe this central point meets your desire to have indicators which can lead to strategies for improvement.

Before reporting NAEP (or other tests) on a local school district level or below, we strongly believe more experience will be required to determine how the national and state program will operate and be used, and what could be specifically envisioned for local progress reports—including the potential impact on curriculum, classroom innovation, and community diversity. Hence, we fully support the Panel's decision not to report locally for the foreseeable future.

We greatly appreciate your efforts and sensitivity to local school boards concerns. If your office wishes to obtain additional information concerning NSBA's position, please contact Associate Executive Director, Michael A. Resnick.

Very truly yours,

ARLENE R. PENFIELD,
President

THOMAS A. SHANNON,
Executive Director

PREPARED STATEMENT OF WILLIAM H. KOLBERG, PRESIDENT OF THE NATIONAL ALLIANCE OF BUSINESS

(The National Alliance of Business is an independent, business-led, nonprofit corporation whose mission is to increase private sector training and job opportunities for economically disadvantaged and long-term unemployed individuals by building and strengthening public/private partnerships of business, government, labor, education, and community-based groups.)

I appreciate the opportunity to submit this statement for consideration by members of subcommittee, and as part of today's hearing to provide a business perspective on the need for federal leadership in research and development to meet the goals of education excellence in this country.

LEADERSHIP THROUGH RESEARCH AND DEVELOPMENT

The focus of the hearing today is one component of the federal role—but a very critical part—research and development. If there is any clear cut role for the federal government, it is in research and development. The Office of Educational Re-

search and Improvement (OERI), and its forerunner the National Institute of Education, was established within the Department of Education to design and implement the education research and development agenda.

My understanding is that much of the work of OERI has been developed over time in response to various congressional mandates and is not, on the whole, a leadership strategy. This hearing is a timely one, because it provides the subcommittee with a chance to rethink the various authorizations of OERI.

There is a critical need to define and focus this role, especially with the setting of national goals. Not that the national goals should determine the entire focus of OERI activities, but they should have a significant influence on its research and development agenda for changing our educational system to meet the goals. We should carefully, but pro-actively, develop the national capacity to provide the knowledge and basic data from which the entire educational system can draw.

My sense is that the OERI research agenda has been set incrementally—step by step—over the years. This has led to funding a broad range of discrete research projects which have not added up to a strategic plan or vision. The result is a Department of Education which is viewed as not providing ample information that the States and local schools can use in their efforts to restructure, and developing information that is not widely or systematically disseminated when completed. We are missing an opportunity for the federal government to provide the leadership in research and to be a resource on education restructuring and national goals.

We need an aggressive and focused research and development agenda to measurably change education quality by the year 2000. It must point the way to real systemic change rather than marginal improvements. It should provide the tools for innovation, experimentation, and information sharing which can have broad application in state and local systems. To have real impact, the research and information must be presented in practical, usable form and for audiences other than educators. This agenda needs to be coupled with other education funds and programs to encourage and facilitate the implementation of what works.

The effort to make systemic change is a joint venture which draws resources, skills, and experiences from all segments of society. If we are to be successful, everyone needs to understand what works and what doesn't. That can only be accomplished through information widely available and written in a manner that is simple and succinct.

Although reaching our education goals should be the primary focus, the research and development agenda should not be limited to those topical areas only. It will take a much more systematic approach to improve education than just those areas identified by the national goals.

The broader purpose should be to have an approach to educational research which reflects an integrated human resources policy. One function of OERI must be to address the whole problem of education reform, from integration of social and human services in elementary and secondary education to institutional restructuring, instruction, curriculum, assessment, and management of the educational process.

The Department needs to be an important catalyst for change, and should have the funds and authority to underwrite much of the developmental work that is necessary. The comprehensive strategy is a means to facilitate OERI's role as a national resource for restructuring. We at the Alliance recognize the role of Congress is helping shape a mission and direction, but there needs to be flexibility in R&D to help find strategies and practices that meet this nation's education needs. We would only caution the Congress against micro-managing the research agenda. The agenda must remain broad and dynamic, but still have a sense of purpose. All of the activities as they are constructed must be interrelated with one goal in mind—a better education system for all children, students, and society.

I see OERI being involved in such activities as:

—Collecting and disseminating data;
—Underwriting the development of assessment instruments, techniques, tools and strategies;
—Evaluating what we are doing in current programs and how well these fit into a restructuring strategy;
—Funding of new and innovative practices and strategies in the form of research and demonstration projects—the projects must offer opportunities for new partners, like business, to participate in these endeavors;
—Disseminating information and materials on "what works" and best practices that are understandable and can be used by audiences other than educators;
—Facilitating the on-going dialogue between researchers, practitioners, and organization, community and business leaders;

—Sponsoring research on instructional and curricula practices which has greater practical application than in the past; and
—Coordinating the federal government's education R&D function to avoid duplication and building a comprehensive strategy.

One of the lead responsibilities should be the development of national educational standards and the methods of assessing individual students against those standards. Since this particular issue has received a lot of attention recently in other congressional hearings, I would like to add a few observations about the OERI role.

The National Alliance of Business and other organizations endorse the notion of establishing a framework for national student assessments. Students would be assessed over a period of years. Cumulative assessment provides multiple opportunities for success rather than a single high-stakes moment of possible failure. It is a system based on a set of national standards.

The assessment system can be helped along substantially through OERI's research and development authority. OERI, more than any other part of the Education Department has the resources and the capacity to direct this effort.

Already, the Assistant Secretary and his staff have been working closely with the National Education Goals Panel. But, beyond working together, OERI must be given the authority and funds for ongoing development of authentic assessment strategies and tools that will indicate where a student is in relation to the standards. They need to build a research agenda that provides alternatives to give us the best choice of assessment tools to achieve the goals.

Mr. Chairman, this is a complex agenda for change. Despite what seem like insurmountable obstacles, a growing current of public opinion demands change in education. All Americans must play a part. We in business are preparing ourselves to play an important role in achieving significant change and improvement.

Our long term agenda, through the Alliance's Center for Excellence in Education and with our partners in the Business Coalition for Education Reform, is to find and implement more effective ways for business involvement.

This education reform effort requires strong national leadership from the federal government in setting the vision and the goals, fostering change, and in ensuring that all the stakeholders carry out their appropriate roles.

THE COLLEGE BOARD
New York, NY, June 21, 1991.

The Honorable CLAIBORNE PELL
U.S. Senate
335 Senate Russell Office Building
Washington, DC.

DEAR SENATOR PELL: I very much appreciated the opportunity to testify last week before your subcommittee on the challenges facing the country in advancing the national education goals and designing appropriate assessment systems.

In addition to my written statement, I would like to submit the enclosed two items for the hearing record:

A copy of the "green book," *Academic Preparation for College: What Students Need to Know and Be Able to Do.* The College Board developed and issued this set of standards after extensive consensus building with secondary and postsecondary educators across the country.

The latest annual report on the Advanced Placement (AP) program. AP seems to embody much of what the current testing movement is calling for—testing in subject areas that are part of the school curriculum, a mix of performance-based measures along with multiple choice test items, and teachers involved in the development of course content and test construction.

I hope these materials are useful to the subcommittee.

Please call on me or Larry Gladieux of our Washington Office whenever the College Board can help.

Sincerely,

DONALD M. STEWART
President

(Due to the high cost of printing, the documents referred to is retained in the files of the committee.)

This wraps up this hearing, and the subcommittee is adjourned.
[Whereupon, at 12:15 p.m., the subcommittee was adjourned.]

○

ISBN 0-16-036815-4

90000

9 780160 368158